

Springer:
COMPLEXITY

Eric Bertin

Statistical Physics of Complex Systems

A Concise Introduction

Second Edition

 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.

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, School of Administrative Studies, York University, 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

Eric Bertin

Statistical Physics of Complex Systems

A Concise Introduction

Second Edition

 Springer

Eric Bertin
LIPhy, CNRS and Université Grenoble
Alpes
Grenoble
France

ISBN 978-3-319-42338-8 ISBN 978-3-319-42340-1 (eBook)
DOI 10.1007/978-3-319-42340-1

Library of Congress Control Number: 2016944901

1st edition: © The Author(s) 2012

2nd edition: © 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

Preface to the Second Edition

The first edition of this book was written on purpose in a very concise, booklet format, to make it easily accessible to a broad interdisciplinary readership of science students and research scientists with an interest in the theoretical modeling of complex systems. Readers were assumed to typically have some bachelor level background in mathematical methods, but no a priori knowledge in statistical physics.

A few years after this first edition, it has appeared relevant to significantly expand it to a full—though still relatively concise—book format in order to include a number of important topics that were not covered in the first edition, thereby raising the number of chapters from three to six. These new topics include non-conserved particles, evolutionary population dynamics, networks (Chap. 4), properties of both individual and coupled simple dynamical systems (Chap. 5), as well as probabilistic issues like convergence theorems for the sum and the extreme values of a large set of random variables (Chap. 6). A few short appendices have also been included, notably to give some technical hints on how to perform simple stochastic simulations in practice.

In addition to these new chapters, the first three chapters have also been significantly updated. In Chap. 1, the discussions of phase transitions and of disordered systems have been slightly expanded. The most important changes in these previously existing chapters concern Chap. 2. The Langevin and Fokker–Planck equations are now presented in separate subsections, including brief discussions about the case of multiplicative noise, the case of more than one degree of freedom, and the Kramers–Moyal expansion. The discussion of anomalous diffusion now focuses on heuristic arguments, while the presentation of the Generalized Central Limit Theorem has been postponed to Chap. 6. Chapter 2 then ends with a discussion of several aspects of the relaxation to equilibrium. Finally, Chap. 3 has also undergone some changes, since the presentation of the Kuramoto model has been deferred to Chap. 5, in the context of deterministic systems. The remaining material of Chap. 3 has then been expanded, with discussions of the Schelling model with

two types of agents, of the dissipative Zero Range Process, and of assemblies of active particles with nematic symmetries.

Although the size of this second edition is more than twice the size of the first one, I have tried to keep the original spirit of the book, so that it could remain accessible to a broad, non-specialized, readership. The presentations of all topics are limited to concise introductions, and are kept to a relatively elementary level—not avoiding mathematics, though. The reader interested in learning more on a specific topic is then invited to look at other sources, like specialized monographs or review articles.

Grenoble, France
May 2016

Eric Bertin

Preface to the First Edition

In recent years, statistical physics started raising the interest of a broad community of researchers in the field of complex system sciences, ranging from biology to social sciences, economics or computer sciences. More generally, a growing number of graduate students and researchers feel the need for learning some basics concepts and questions coming from other disciplines, leading for instance to the organization of recurrent interdisciplinary summer schools.

The present booklet is partly based on the introductory lecture on statistical physics given at the French Summer School on Complex Systems held both in Lyon and Paris during the summers 2008 and 2009, and jointly organized by two French Complex Systems Institutes, the “Institut des Systèmes Complexes Paris Ile de France” (ISC-PIF) and the “Institut Rhône-Alpin des Systèmes Complexes” (IXXI). This introductory lecture was aimed at providing the participants with a basic knowledge of the concepts and methods of statistical physics so that they could later on follow more advanced lectures on diverse topics in the field of complex systems. The lecture has been further extended in the framework of the second year of Master in “Complex Systems Modelling” of the Ecole Normale Supérieure de Lyon and Université Lyon 1, whose courses take place at IXXI.

It is a pleasure to thank Guillaume Beslon, Tommaso Roscilde and Sébastien Grauwin, who were also involved in some of the lectures mentioned above, as well as Pablo Jensen for his efforts in setting up an interdisciplinary Master course on complex systems, and for the fruitful collaboration we had over the last years.

Lyon, France
June 2011

Eric Bertin

Contents

1	Equilibrium Statistical Physics	1
1.1	Microscopic Dynamics of a Physical System	1
1.1.1	Conservative Dynamics	1
1.1.2	Properties of the Hamiltonian Formulation	3
1.1.3	Many-Particle System	5
1.1.4	Case of Discrete Variables: Spin Models	6
1.2	Statistical Description of an Isolated System at Equilibrium	6
1.2.1	Notion of Statistical Description: A Toy Model	6
1.2.2	Fundamental Postulate of Equilibrium Statistical Physics	7
1.2.3	Computation of $\Omega(E)$ and $S(E)$: Some Simple Examples	8
1.2.4	Distribution of Energy Over Subsystems and Statistical Temperature	10
1.3	Equilibrium System in Contact with Its Environment	12
1.3.1	Exchanges of Energy	12
1.3.2	Canonical Entropy	16
1.3.3	Exchanges of Particles with a Reservoir: The Grand-Canonical Ensemble	17
1.4	Phase Transitions and Ising Model	18
1.4.1	Ising Model in Fully Connected Geometry	19
1.4.2	Ising Model with Finite Connectivity	21
1.4.3	Renormalization Group Approach	23
1.5	Disordered Systems and Glass Transition	29
1.5.1	Theoretical Spin-Glass Models	29
1.5.2	A Toy Model for Spin-Glasses: The Mattis Model	30
1.5.3	The Random Energy Model	32
	References	35

2	Non-stationary Dynamics and Stochastic Formalism	37
2.1	Markovian Stochastic Processes and Master Equation	38
2.1.1	Definition of Markovian Stochastic Processes	38
2.1.2	Master Equation and Detailed Balance	39
2.1.3	A Simple Example: The One-Dimensional Random Walk	41
2.2	Langevin Equation	44
2.2.1	Phenomenological Approach	44
2.2.2	Basic Properties of the Linear Langevin Equation	46
2.2.3	More General Forms of the Langevin Equation	49
2.2.4	Relation to Random Walks	51
2.3	Fokker-Planck Equation	53
2.3.1	Continuous Limit of a Discrete Master Equation	53
2.3.2	Kramers-Moyal Expansion	54
2.3.3	More General Forms of the Fokker-Planck Equation	55
2.4	Anomalous Diffusion: Scaling Arguments	57
2.4.1	Importance of the Largest Events	57
2.4.2	Superdiffusive Random Walks	59
2.4.3	Subdiffusive Random Walks	61
2.5	Fast and Slow Relaxation to Equilibrium	63
2.5.1	Relaxation to Canonical Equilibrium	63
2.5.2	Dynamical Increase of the Entropy	65
2.5.3	Slow Relaxation and Physical Aging	67
	References	71
3	Statistical Physics of Interacting Macroscopic Units	73
3.1	Dynamics of Residential Moves	74
3.1.1	A Simplified Version of the Schelling Model	75
3.1.2	Condition for Phase Separation	77
3.1.3	The ‘True’ Schelling Model: Two Types of Agents	80
3.2	Driven Particles on a Lattice: Zero-Range Process	81
3.2.1	Definition and Exact Steady-State Solution	81
3.2.2	Maximal Density and Condensation Phenomenon	82
3.2.3	Dissipative Zero-Range Process	83
3.3	Collective Motion of Active Particles	86
3.3.1	Derivation of Continuous Equations	87
3.3.2	Phase Diagram and Instabilities	91
3.3.3	Varying the Symmetries of Particles	92
	References	94
4	Beyond Assemblies of Stable Units	95
4.1	Non-conserved Particles: Reaction-Diffusion Processes	95
4.1.1	Mean-Field Approach of Absorbing Phase Transitions	96
4.1.2	Fluctuations in a Fully Connected Model	98

- 4.2 Evolutionary Dynamics 101
 - 4.2.1 Statistical Physics Modeling of Evolution in Biology. 101
 - 4.2.2 Selection Dynamics Without Mutation 103
 - 4.2.3 Quasistatic Evolution Under Mutation 105
- 4.3 Dynamics of Networks. 109
 - 4.3.1 Random Networks. 110
 - 4.3.2 Small-World Networks 112
 - 4.3.3 Preferential Attachment 114
- References 116
- 5 Statistical Description of Deterministic Systems 119**
 - 5.1 Basic Notions on Deterministic Systems. 119
 - 5.1.1 Fixed Points and Simple Attractors 119
 - 5.1.2 Bifurcations 121
 - 5.1.3 Chaotic Dynamics. 123
 - 5.2 Deterministic Versus Stochastic Dynamics 125
 - 5.2.1 Qualitative Differences and Similarities 125
 - 5.2.2 Stochastic Coarse-Grained Description
of a Chaotic Map 127
 - 5.2.3 Statistical Description of Chaotic Systems 128
 - 5.3 Globally Coupled Dynamical Systems 130
 - 5.3.1 Coupling Low-Dimensional Dynamical Systems 130
 - 5.3.2 Description in Terms of Global Order Parameters 131
 - 5.3.3 Stability of the Fixed Point of the Global System 132
 - 5.4 Synchronization Transition 135
 - 5.4.1 The Kuramoto Model of Coupled Oscillators 135
 - 5.4.2 Synchronized Steady State 137
 - References 139
- 6 A Probabilistic Viewpoint on Fluctuations and Rare Events. 141**
 - 6.1 Global Fluctuations as a Random Sum Problem 141
 - 6.1.1 Law of Large Numbers and Central Limit Theorem. 141
 - 6.1.2 Generalization to Variables with Infinite Variances 143
 - 6.1.3 Case of Non-identically Distributed Variables. 146
 - 6.1.4 Case of Correlated Variables 149
 - 6.1.5 Coarse-Graining Procedures and Law
of Large Numbers. 151
 - 6.2 Rare and Extreme Events 153
 - 6.2.1 Different Types of Rare Events. 153
 - 6.2.2 Extreme Value Statistics 154
 - 6.2.3 Statistics of Records 156

- 6.3 Large Deviation Functions 158
 - 6.3.1 A Simple Example: The Ising Model
in a Magnetic Field 158
 - 6.3.2 Explicit Computations of Large Deviation Functions 160
 - 6.3.3 A Natural Framework to Formulate Statistical Physics 161
- References 162
- Appendix A: Dirac Distribution.** 165
- Appendix B: Numerical Simulations of Markovian
Stochastic Processes** 167
- Appendix C: Drawing Random Variables with Prescribed
Distributions.** 169

Introduction

Generally speaking, the goals of statistical physics can be summarized as follows: on the one hand to study systems composed of a large number of interacting ‘units’, and on the other hand to predict the macroscopic (or collective) behavior of the system considered from the microscopic laws ruling the dynamics of the individual ‘units’. These two goals are, to some extent, also shared by what is nowadays called ‘complex systems science’. However, the specificity of statistical physics is that:

- The ‘units’ considered are in most cases atoms or molecules, for which the individual microscopic laws are known from fundamental physical theories—at variance with other fields like social sciences for example, where little is known about the quantitative behavior of individuals.
- These atoms, or molecules, are often all of the same type, or at most of a few different types—in contrast to biological or social systems for instance, where the individual ‘units’ may all differ, or at least belong to a large number of different types.

For these reasons, systems studied in the framework of statistical physics may be considered as among the simplest examples of complex systems. One further specificity of statistical physics with respect to other sciences aiming at describing the collective behavior of complex systems is that it allows for a rather well-developed mathematical treatment.

The present book is divided into six chapters. Chapter 1 deals with equilibrium statistical physics, trying to expose in a concise way the main concepts of this theory, and paying specific attention to those concepts that could be more generally relevant to complex system sciences. Of particular interest is on the one hand the phenomenon of phase transition, and on the other hand the study of disordered systems. Chapter 2 mainly aims at describing dynamical effects like diffusion or relaxation, in the framework of Markovian stochastic processes. A simple description of the formalism is provided, together with a discussion of random walk processes, as well as Langevin and Fokker–Planck equations. Anomalous diffusion processes are also briefly described, as well as some generic properties of the relaxation of stochastic processes to equilibrium.

Chapter 3 deals with the generic issue of the statistical description of large systems of interacting ‘units’ under nonequilibrium conditions. These nonequilibrium units may be for instance particles driven by an external field, social agents moving from one flat to another in a city, or self-propelled particles representing in a schematic way bacteria or self-driven colloids. Their description relies on the adaptation of different techniques borrowed from standard statistical physics, including mappings to effective equilibrium systems, Boltzmann approaches (a technique early developed in statistical physics to characterize the dynamics of gases) for systems interacting through binary collisions, or exact solutions when available.

Chapter 4 aims at going beyond the case of stable interacting units, by investigating several possible extensions. The first one is the case of reaction-diffusion processes, in which particles can be created and annihilated, leading to a peculiar type of phase transitions called absorbing phase transitions. The case of population dynamics, in connection with the process of biological evolution, is also presented. The chapter ends with a brief presentation of the statistics of random networks.

After these three chapters dedicated to stochastic processes, Chap. 5 presents some elementary notions on dynamical systems, concerning in particular the fixed points and their stability, the more general concept of attractor, as well as the notion of bifurcation. A discussion on the comparison between deterministic and stochastic dynamics is provided, in connection with coarse-graining issues. Then, the case of globally coupled population of low-dimensional dynamical systems is investigated through the analysis of two different cases, the restabilization of unstable fixed points by the coupling and the synchronization transition in the Kuramoto model of coupled oscillators.

Finally, Chap. 6 presents some basic results of probability theory which are of high interest in a statistical physics context. This chapter deals in particular with the statistics of sums of random variables (Law of Large Numbers, standard and generalized Central Limit Theorems), the statistics of extreme values and records, and the statistics of very rare events as described by the large deviation formalism.

Chapter 1

Equilibrium Statistical Physics

Systems composed of many particles involve a very large number of degrees of freedom, and it is most often uninteresting—not to say hopeless—to try to describe in a detailed way the microscopic state of the system. The aim of statistical physics is rather to restrict the description of the system to a few relevant macroscopic observables, and to predict the average values of these observables, or the relations between them. A standard formalism, called "equilibrium statistical physics", has been developed for systems of physical particles having reached a statistical steady state in the absence of external driving (like heat flux or shearing forces for instance).

In this first part, we shall discuss some of the fundamentals of equilibrium statistical physics. Sect. 1.1 describes the elementary mechanical notions necessary to describe a system of physical particles. Section 1.2 introduces the basic statistical notions and fundamental postulates required to describe in a statistical way a system that exchanges no energy with its environment. The effect of the environment is then taken into account in Sect. 1.3, in the case where the environment does not generate any sustained energy flux in the system. Applications of this general formalism to the description of collective phenomena and phase transitions are presented in Sect. 1.4. Finally, the influence of disorder and heterogeneities, which are relevant in physical systems, but are also expected to play an essential role in many other types of complex systems, is briefly discussed in Sect. 1.5. For further reading on these topics related to equilibrium statistical physics (especially for Sects. 1.2–1.4), we refer the reader to standard textbooks, like e.g. Refs. [1–4].

1.1 Microscopic Dynamics of a Physical System

1.1.1 *Conservative Dynamics*

In the framework of statistical physics, an important type of dynamics is the so-called conservative dynamics in which energy is conserved, meaning that

friction forces are absent, or can be neglected. As an elementary example, consider a particle constrained to move on a one-dimensional horizontal axis x , and attached to a spring, the latter being pinned to a rigid wall. We consider the position $x(t)$ of the particle at time t , as well as its velocity $v(t)$. The force F exerted by the spring on the particle is given by

$$F = -k(x - x_0), \quad (1.1)$$

where x_0 corresponds to the position of repose of the particle, for which the force vanishes. For convenience, we shall in the following choose the origin of the x axis such that $x_0 = 0$.

From the basic laws of classical mechanics, the motion of the particle is described by the evolution equation:

$$m \frac{dv}{dt} = F \quad (1.2)$$

where m is the mass of the particle. We have neglected all friction forces, so that the force exerted by the spring is the only horizontal force (the gravity force, as well as the reaction force exerted by the support, do not have horizontal components in the absence of friction). In terms of x variable, the equation of motion (1.2) reads

$$m \frac{d^2x}{dt^2} = -kx. \quad (1.3)$$

The generic solution of this equation is

$$x(t) = A \cos(\omega t + \phi), \quad \omega = \sqrt{\frac{k}{m}}. \quad (1.4)$$

The constants A and ϕ are determined by the initial conditions, namely the position and velocity at time $t = 0$.

The above dynamics can be reformulated in the so-called Hamiltonian formalism. Let us introduce the momentum $p = mv$, and the kinetic energy $E_c = \frac{1}{2}mv^2$. In terms of momentum, the kinetic energy reads $E_c = p^2/2m$. The potential energy U of the spring, defined by $F = -dU/dx$, is given by $U = \frac{1}{2}kx^2$. The Hamiltonian $H(x, p)$ is defined as

$$H(x, p) = E_c(p) + U(x). \quad (1.5)$$

In the present case, this definition yields

$$H(x, p) = \frac{p^2}{2m} + \frac{1}{2}kx^2. \quad (1.6)$$

In the Hamiltonian formulation, the equations of motion read¹

¹For a more detailed introduction to the Hamiltonian formalism, see, e.g., Ref. [5].

$$\frac{dx}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial x}. \quad (1.7)$$

On the example of the particle attached to a spring, these equations give

$$\frac{dx}{dt} = \frac{p}{m}, \quad \frac{dp}{dt} = -kx, \quad (1.8)$$

from which one recovers Eq. (1.3) by eliminating p . Hence it is seen on the above example that the Hamiltonian formalism is equivalent to the standard law of motion (1.2).

1.1.2 Properties of the Hamiltonian Formulation

Energy conservation. The Hamiltonian formulation has interesting properties, namely energy conservation and time-reversal invariance. We define the total energy $E(t)$ as $E(t) = H(x(t), p(t)) = E_c(p(t)) + U(x(t))$. It is easily shown that the total energy is conserved during the evolution of the system²

$$\frac{dE}{dt} = \frac{\partial H}{\partial x} \frac{dx}{dt} + \frac{\partial H}{\partial p} \frac{dp}{dt}. \quad (1.9)$$

Using Eq. (1.7), one has

$$\frac{dE}{dt} = \frac{\partial H}{\partial x} \frac{\partial H}{\partial p} + \frac{\partial H}{\partial p} \left(-\frac{\partial H}{\partial x} \right) = 0, \quad (1.10)$$

so that the energy E is conserved. This is confirmed by a direct calculation on the example of the particle attached to a spring:

$$\begin{aligned} E(t) &= \frac{p(t)^2}{2m} + \frac{1}{2}kx(t)^2 \\ &= \frac{1}{2m}m^2\omega^2 A^2 \sin^2(\omega t + \phi) + \frac{1}{2}kA^2 \cos^2(\omega t + \phi). \end{aligned} \quad (1.11)$$

²The concept of energy, introduced here on a specific example, plays a fundamental role in physics. Although any precise definition of the energy is necessarily formal and abstract, the notion of energy can be thought of intuitively as a quantity that can take very different forms (kinetic, electromagnetic or gravitational energy, but also internal energy exchanged through heat transfers) in such a way that the total amount of energy remains constant. Hence an important issue is to describe how energy is transferred from one form to another. For instance, in the case of the particle attached to a spring, the kinetic energy E_c and potential energy U of the spring are continuously exchanged, in a reversible manner. In the presence of friction forces, kinetic energy would also be progressively converted, in an irreversible way, into internal energy, thus raising the temperature of the system.

Given that $\omega^2 = k/m$, one finds

$$E(t) = \frac{1}{2}kA^2 (\sin^2(\omega t + \phi) + \cos^2(\omega t + \phi)) = \frac{1}{2}kA^2 \quad (1.12)$$

which is indeed a constant.

Time reversal invariance. Another important property of the Hamiltonian dynamics is its time reversibility. To illustrate the meaning of time reversibility, let us imagine that we film the motion of the particle with a camera, and that we project it backward. If the backward motion is also a possible motion, meaning that nothing is unphysical in the backward projected movie, then the equations of motion are time-reversible.

More formally, we consider the trajectory $x(t)$, $0 \leq t \leq t_0$, and define the reversed time $t' = t_0 - t$. Starting from the equations of motion (1.7) expressed with t , x and p , time reversal is implemented by replacing t with $t_0 - t'$, x with x' and p with $-p'$, yielding

$$-\frac{dx}{dt'} = -\frac{\partial H}{\partial p'}, \quad \frac{dp'}{dt'} = -\frac{\partial H}{\partial x'}. \quad (1.13)$$

Changing the overall sign in the first equation, one recovers Eq. (1.7) for the primed variables, meaning that the time-reversed trajectory is also a physical trajectory.

Note that time-reversibility holds only as long as friction forces are neglected. The latter break time reversal invariance, and this explains why our everyday-life experience seems to contradict time reversal invariance. For instance, when a glass falls down onto the floor and breaks into pieces, it is hard to believe that the reverse trajectory, in which pieces would come together and the glass would jump onto the table, is also a possible trajectory, as nobody has ever seen this phenomenon occur. In order to reconcile macroscopic irreversibility and microscopic reversibility of trajectories, the point of view of statistical physics is to consider that the reverse trajectory is possible, but has a very small probability to occur as only very few initial conditions could lead to this trajectory. So in practice, the corresponding trajectory is never observed.

Phase-space representation. Finally, let us mention that it is often convenient to consider the Hamiltonian dynamics as occurring in an abstract space called ‘phase space’ rather than in real space. Physical space is described in the above example by the coordinate x . The equations of motion (1.7) allow the position x and momentum p of the particle to be determined at any time once the initial position and momentum are known. So it is interesting to introduce an abstract representation space containing both position and momentum. In this example, it is a two-dimensional space, but it could be of higher dimension in more general situations. This representation space is often called “phase space”. For the particle attached to the spring, the trajectories in this phase space are ellipses. Rescaling the coordinates in an appropriate way, one can transform the ellipse into a circle, and the energy can be identified with the square of the radius of the circle. To illustrate this property, let us define the new phase-space coordinates X and Y as

$$X = \sqrt{\frac{k}{2}}x, \quad Y = \frac{p}{\sqrt{2m}}. \quad (1.14)$$

Then the energy E can be written as

$$E = \frac{p^2}{2m} + \frac{1}{2}kx^2 = X^2 + Y^2. \quad (1.15)$$

As the energy is fixed, the trajectory of the particle is a circle of radius \sqrt{E} in the (X, Y) -plane.

1.1.3 Many-Particle System

In a more general situation, a physical system is composed of N particles in a 3-dimensional space. The position of particle i is described by a vector \mathbf{x}_i , and its velocity by \mathbf{v}_i , $i = 1, \dots, N$. In the Hamiltonian formalism, it is often convenient to introduce generalized coordinates q_j and momenta p_j which are scalar quantities, with $j = 1, \dots, 3N$: (q_1, q_2, q_3) are the components of the vector \mathbf{x}_1 describing the position of particle 1, (q_4, q_5, q_6) are the component of \mathbf{x}_2 , and so on. Similarly, (p_1, p_2, p_3) are the components of the momentum vector $m\mathbf{v}_1$ of particle 1, (p_4, p_5, p_6) are the components of $m\mathbf{v}_2$, etc. With these notations, the Hamiltonian of the N -particle system is defined as

$$H(q_1, \dots, q_{3N}, p_1, \dots, p_{3N}) = \sum_{j=1}^{3N} \frac{p_j^2}{2m} + U(q_1, \dots, q_{3N}). \quad (1.16)$$

The first term in the Hamiltonian is the kinetic energy, and the last one is the potential (or interaction) energy. The equations of motion read

$$\frac{dq_j}{dt} = \frac{\partial H}{\partial p_j}, \quad \frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j}, \quad j = 1, \dots, 3N. \quad (1.17)$$

The properties of energy conservation and time-reversal invariance also hold in this more general formulation, and are derived in the same way as above. As an illustration, typical examples of interaction energy U include

- $U = 0$: case of free particles.
- $U = -\sum_{i=1}^N \mathbf{h}_i \mathbf{x}_i$: particles interacting with an external field, for instance the gravity field, or an electric field.
- $U = \sum_{i \neq i'} V(\mathbf{x}_i - \mathbf{x}_{i'})$: pair interaction potential.

1.1.4 Case of Discrete Variables: Spin Models

As a simplified picture, a spin may be thought of as a magnetization \mathbf{S} associated to an atom. The dynamics of spins is ruled by quantum mechanics (the theory that governs particles at the atomic scale), which is outside the scope of the present book. However, in some situations, the configuration of a spin system can be represented in a simplified way as a set of binary “spin variables” $s_i = \pm 1$, and the corresponding energy takes the form

$$E = -J \sum_{\langle i,j \rangle} s_i s_j - h \sum_{i=1}^N s_i. \quad (1.18)$$

The parameter J is the coupling constant between spins, while h is the external magnetic field. The first sum corresponds to a sum over nearest neighbor sites on a lattice, but other types of interaction could be considered. This model is called the Ising model. It provides a qualitative description of the phenomenon of ferromagnetism observed in metals like iron, in which a spontaneous macroscopic magnetization appears below a certain critical temperature. In addition, the Ising model turns out to be very useful to illustrate some important concepts of statistical physics.

In what follows, we shall consider the words “energy” and “Hamiltonian” as synonyms, and the corresponding notations E and H as equivalent.

1.2 Statistical Description of an Isolated System at Equilibrium

1.2.1 Notion of Statistical Description: A Toy Model

Let us consider a toy model in which a particle is moving on a ring with L sites. Time is discretized, meaning that for instance every second the particle moves to the next site. The motion is purely deterministic: given the position at time $t = 0$, one can compute the position $i(t)$ at any later time. Now assume that there is an observable ε_i on each site i . It could be for instance the height of the site, or any arbitrary observable that characterizes the state of the particle when it is at site i .

A natural question would be to know what the average value

$$\langle \varepsilon \rangle = \frac{1}{T} \sum_{t=1}^T \varepsilon_{i(t)} \quad (1.19)$$

is after a large observation time T . Two different approaches to this question can be proposed:

- Simulate the dynamics of the model on a computer, and measure directly $\langle \varepsilon \rangle$.

- Use the concept of probability as a shortcut, and write

$$\langle \varepsilon \rangle = \sum_{i=1}^L P_i \varepsilon_i \quad (1.20)$$

where the probability P_i to be on site i is defined as

$$P_i = \frac{\text{time spent on site } i}{\text{total time } T}, \quad (1.21)$$

namely the fraction of time spent on site i .

The probability P_i can be calculated or measured by simulating the dynamics, but it can also be estimated directly: if the particle has turned a lot of times around the ring, the fraction of time spent on each site is the same, $P_i = 1/L$. Hence all positions of the particle are equiprobable, and the average value $\langle \varepsilon \rangle$ is obtained as a flat average over all sites. Of course, more complicated situations may occur, and the concept of probability remains useful beyond the simple equiprobability situation described above.

1.2.2 *Fondamental Postulate of Equilibrium Statistical Physics*

We consider a physical system composed of N particles. The microscopic configuration of the system is described by $(\mathbf{x}_i, \mathbf{p}_i = m\mathbf{v}_i)$, $i = 1, \dots, N$, or $s_i = \pm 1$, $i = 1, \dots, N$, for spin systems.

The total energy E of the system, given for instance for systems of identical particles by

$$E = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + U(\mathbf{x}_1, \dots, \mathbf{x}_N), \quad (1.22)$$

or for spins systems by

$$E = -J \sum_{(i,j)} s_i s_j - h \sum_{i=1}^N s_i, \quad (1.23)$$

is constant as a function of time (or may vary within a tiny interval $[E, E + \delta E]$, in particular for spin systems). Accordingly, starting from an initial condition with energy E , the system can only visit configurations with the same energy. In the absence of further information, it is legitimate to postulate that all configurations with the same energy as the initial one have the same probability to be visited. This leads us to the **fundamental postulate of equilibrium statistical physics**:

Given an energy E , all configurations with energy E have equal nonzero probabilities. Other configurations have zero probability.

The corresponding probability distribution is called the microcanonical distribution or microcanonical ensemble for historical reasons (a probability distribution can be thought of as describing an infinite set of copies—an ensemble—of a given system).

A quantity that plays an important role is the “volume” $\Omega(E)$ occupied in phase-space by all configurations with energy E . For systems with continuous degrees of freedom, $\Omega(E)$ is the area of the hypersurface defined by fixing the energy E . For systems with discrete configurations (spins), $\Omega(E)$ is the number of configurations with energy E . The Boltzmann entropy is defined as

$$S(E) = k_B \ln \Omega(E), \quad (1.24)$$

where $k_B = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant. This constant has been introduced both for historical and practical reasons, but from a theoretical viewpoint, its specific value plays no role, so that we shall set it to $k_B = 1$ in the following (this could be done for instance by working with specific units of temperature and energy such that $k_B = 1$ in these units).

The notion of entropy is a cornerstone of statistical physics. First introduced in the context of thermodynamics (the theory of the balance between mechanical energy transfers and heat exchanges), entropy was later on given a microscopic interpretation in the framework of statistical physics. Basically, entropy is a measure of the number of available microscopic configurations compatible with the macroscopic constraints. More intuitively, entropy can be interpreted as a measure of ‘disorder’ (disordered macroscopic states often correspond to a larger number of microscopic configurations than macroscopically ordered states), though the correspondence between the two notions is not necessarily straightforward and may fail in some cases like in the liquid-solid transition of hard spheres. Another popular interpretation, in relation to information theory, is to consider entropy as a measure of the lack of information on the system: the larger the number of accessible microscopic configurations, the less information is available on the system (in an extreme case, if the system can be with equal probability in any microscopic configuration, one has no information on the state of the system).

Let us now give a few simple examples of computation of the entropy.

1.2.3 Computation of $\Omega(E)$ and $S(E)$: Some Simple Examples

Paramagnetic spin model. We consider a model of independent spins, interacting only with a uniform external field. The corresponding energy is given by

$$E = -h \sum_{i=1}^N s_i, \quad s_i = \pm 1. \quad (1.25)$$

The phase space (or here simply configuration space) is given by the list of values (s_1, \dots, s_N) . The question is to know how many configurations there are with a given energy E . In this specific example, it is easily seen that fixing the energy E amounts to fixing the magnetization $M = \sum_{i=1}^N s_i$. Let us denote as N_+ the number of spins with value $+1$ ('up' spins). The magnetization is given by $M = N_+ - (N - N_+) = 2N_+ - N$, so that fixing M is in turn equivalent to fixing N_+ . From basic combinatorial arguments, the number of configurations with a given number of 'up' spins is given by

$$\Omega = \frac{N!}{N_+!(N - N_+)!}. \quad (1.26)$$

Using the relation

$$N_+ = \frac{1}{2} \left(N - \frac{E}{h} \right), \quad (1.27)$$

one can express Ω as a function of E :

$$\Omega(E) = \frac{N!}{\left[\frac{1}{2}(N - E/h) \right]! \left[\frac{1}{2}(N + E/h) \right]!}. \quad (1.28)$$

The entropy $S(E)$ is given by

$$\begin{aligned} S(E) &= \ln \Omega(E) \\ &= \ln N! - \ln \left[\frac{1}{2} \left(N - \frac{E}{h} \right) \right]! - \ln \left[\frac{1}{2} \left(N + \frac{E}{h} \right) \right]! \end{aligned} \quad (1.29)$$

Using Stirling's approximation, valid for large N

$$\ln N! \approx N \ln N - N, \quad (1.30)$$

one finds

$$S(E) = N \ln N - \frac{N + E/h}{2} \ln \frac{N + E/h}{2} - \frac{N - E/h}{2} \ln \frac{N - E/h}{2}. \quad (1.31)$$

Perfect gas of independent particles. As a second example, we consider a gas of independent particles confined into a cubic container of volume $V = L^3$. The generalized coordinates q_j satisfy the constraints

$$0 \leq q_j \leq L, \quad j = 1, \dots, L. \quad (1.32)$$

The energy E comes only from the kinetic contribution:

$$E = \sum_{j=1}^{3N} \frac{p_j^2}{2m}. \quad (1.33)$$

The accessible volume in phase space is the product of the accessible volume for each particle, times the area of the hypersphere of radius $\sqrt{2mE}$, embedded in a $3N$ -dimensional space. The area of the hypersphere of radius R in a D -dimensional space is

$$\mathcal{A}_D(R) = \frac{D\pi^{D/2}}{\Gamma(\frac{D}{2} + 1)} R^{D-1}, \quad (1.34)$$

where $\Gamma(x) = \int_0^\infty dt t^{x-1} e^{-t}$ is the Euler Gamma function (a generalization of the factorial to real values, satisfying $\Gamma(n) = (n-1)!$ for integer $n \geq 1$). Hence the accessible volume $\Omega_V(E)$ is given by

$$\Omega_V(E) = \frac{3N\pi^{3N/2}}{\Gamma(\frac{3N}{2} + 1)} \sqrt{2m}^{3N-1} V^N E^{\frac{3N-1}{2}}. \quad (1.35)$$

The corresponding entropy reads, assuming $N \gg 1$,

$$S_V(E) = \ln \Omega(E) = S_0 + \frac{3N}{2} \ln E + N \ln V \quad (1.36)$$

with

$$S_0 = \ln \left(\frac{3N\pi^{3N/2}}{\Gamma(\frac{3N}{2} + 1)} \sqrt{2m}^{3N} \right). \quad (1.37)$$

Note that in principle, some corrections need to be included to take into account quantum effects, namely the fact that quantum particles are indistinguishable. This allows in particular $\Omega(E)$ to be made dimensionless, thus rendering the entropy independent of the system of units chosen. Quantum effects are also important in order to recover the extensivity of the entropy, that is, the fact that the entropy is proportional to the number N of particles. In the present form, $N \ln N$ terms are present, making the entropy grow faster than the system size. This is related to the so-called Gibbs paradox. However, we shall not describe these effects in more details here, and refer the reader to standard textbooks [1–4].

1.2.4 Distribution of Energy Over Subsystems and Statistical Temperature

Let us consider an isolated system, with fixed energy and number of particles. We then imagine that the system is partitioned into two subsystems \mathcal{S}_1 and \mathcal{S}_2 , the two subsystems being separated by a wall which allows energy exchanges, but not

exchanges of matter. The total energy of the system $E = E_1 + E_2$ is fixed, but the energies E_1 and E_2 fluctuate due to thermal exchanges.

For a fixed energy E , let us evaluate the number $\Omega(E_1|E)$ of configurations of the system such that the energy of \mathcal{S}_1 has a given value E_1 . In the absence of long-range forces in the system, the two subsystems can be considered as statistically independent (apart from the total energy constraint), leading to

$$\Omega(E_1|E) = \Omega_1(E_1)\Omega_2(E - E_1), \quad (1.38)$$

where $\Omega_k(E_k)$ is the number of configurations of \mathcal{S}_k .

The most probable value E_1^* of the energy E_1 maximizes by definition $\Omega(E_1|E)$, or equivalently $\ln \Omega(E_1|E)$:

$$\left. \frac{\partial}{\partial E_1} \ln \Omega(E_1|E) \right|_{E_1^*} = 0. \quad (1.39)$$

Combining Eqs. (1.38) and (1.39), one finds

$$\left. \frac{\partial \ln \Omega_1}{\partial E_1} \right|_{E_1^*} = \left. \frac{\partial \ln \Omega_2}{\partial E_2} \right|_{E_2^* = E - E_1^*}. \quad (1.40)$$

Thus it turns out that two quantities defined independently in each subsystem are equal at equilibrium. Namely, defining

$$\beta_k \equiv \left. \frac{\partial \ln \Omega_k}{\partial E_k} \right|_{E_k^*}, \quad k = 1, 2, \quad (1.41)$$

one has $\beta_1 = \beta_2$. This is the reason why the quantity β_k is called the statistical temperature of \mathcal{S}_k . In addition, it can be shown that for large systems, the common value of β_1 and β_2 is also equal to

$$\beta = \frac{\partial S}{\partial E} \quad (1.42)$$

computed for the global isolated system.

To identify the precise link between β and the standard thermodynamic temperature, we notice that in thermodynamics, one has for a system that exchanges no work with its environment:

$$dE = T dS, \quad (1.43)$$

which indicates that $\beta = 1/T$ (we recall that we have set $k_B = 1$). This is further confirmed on the example of the perfect gas, for which one finds using Eq. (1.36)

$$\beta \equiv \frac{\partial S}{\partial E} = \frac{3N}{2E}, \quad (1.44)$$

or equivalently

$$E = \frac{3N}{2\beta}. \quad (1.45)$$

Besides, one has from the kinetic theory of gases

$$E = \frac{3}{2}NT \quad (1.46)$$

(which is nothing but equipartition of energy), leading again to the identification $\beta = 1/T$. Hence, in the microcanonical ensemble, one generically defines temperature T through the relation

$$\frac{1}{T} = \frac{\partial S}{\partial E}. \quad (1.47)$$

We now further illustrate this relation on the example of the paramagnetic crystal that we already encountered earlier. From Eq. (1.31), one has

$$\frac{1}{T} = \frac{\partial S}{\partial E} = \frac{1}{2h} \ln \frac{N - E/h}{N + E/h}. \quad (1.48)$$

This last equation can be inverted to express the energy E as a function of temperature, yielding

$$E = -Nh \tanh \frac{h}{T}. \quad (1.49)$$

This relation has been obtained by noticing that $x = \tanh y$ is equivalent to

$$y = \frac{1}{2} \ln \left(\frac{1+x}{1-x} \right). \quad (1.50)$$

In addition, from the relation $E = -Mh$, where $M = \sum_{i=1}^N s_i$ is the total magnetization, one obtains as a byproduct

$$M = N \tanh \frac{h}{T}. \quad (1.51)$$

1.3 Equilibrium System in Contact with Its Environment

1.3.1 Exchanges of Energy

Realistic systems are most often not isolated, but they rather exchange energy with their environment. A natural idea is then to describe the system \mathcal{S} of interest as a macroscopic subsystem of a large isolated system $\mathcal{S} \cup \mathcal{R}$, where \mathcal{R} is the environment,

or energy reservoir. The total energy $E_{\text{tot}} = E + E_{\mathcal{R}}$ is fixed. A configuration C_{tot} of the total system can be written as $C_{\text{tot}} = (C, C_{\mathcal{R}})$, where C is a configuration of \mathcal{S} and $C_{\mathcal{R}}$ is a configuration of \mathcal{R} . The total system $\mathcal{S} \cup \mathcal{R}$ is isolated and at equilibrium, so that it can be described within the macrocanonical framework:

$$P_{\text{tot}}(C_{\text{tot}}) = \frac{1}{\Omega_{\text{tot}}(E_{\text{tot}})}, \quad C_{\text{tot}} = (C, C_{\mathcal{R}}). \quad (1.52)$$

To obtain the probability of a configuration C of \mathcal{S} , one needs to sum $P_{\text{tot}}(C_{\text{tot}})$ over all configurations $C_{\mathcal{R}}$ of \mathcal{R} compatible with the total energy E_{tot} , namely

$$P(C) = \sum_{C_{\mathcal{R}}: E_{\mathcal{R}} = E_{\text{tot}} - E(C)} P_{\text{tot}}(C, C_{\mathcal{R}}) = \frac{\Omega_{\mathcal{R}}(E_{\text{tot}} - E(C))}{\Omega_{\text{tot}}(E_{\text{tot}})}. \quad (1.53)$$

We introduce the entropy of the reservoir $S_{\mathcal{R}}(E_{\mathcal{R}}) = \ln \Omega_{\mathcal{R}}(E_{\mathcal{R}})$. Under the assumption that $E(C) \ll E_{\text{tot}}$, one has

$$S_{\mathcal{R}}(E_{\text{tot}} - E(C)) \approx S_{\mathcal{R}}(E_{\text{tot}}) - E(C) \left. \frac{\partial S_{\mathcal{R}}}{\partial E_{\mathcal{R}}} \right|_{E_{\text{tot}}}. \quad (1.54)$$

One also has

$$\left. \frac{\partial S_{\mathcal{R}}}{\partial E_{\mathcal{R}}} \right|_{E_{\text{tot}}} \approx \left. \frac{\partial S_{\mathcal{R}}}{\partial E_{\mathcal{R}}} \right|_{E_{\mathcal{R}}^*} = \frac{1}{T} \quad (1.55)$$

where T is the temperature of the reservoir. Altogether, we have

$$P(C) = \frac{\Omega_{\mathcal{R}}(E_{\text{tot}})}{\Omega_{\text{tot}}(E_{\text{tot}})} e^{-E(C)/T}. \quad (1.56)$$

Note that the prefactor $\Omega_{\mathcal{R}}/\Omega_{\text{tot}}$ depends on the total energy E_{tot} , while we would like $P(C)$ to depend only on the energy E of the system considered. This problem can however be bypassed by noticing that the distribution $P(C)$ should be normalized to unity, namely, $\sum_C P(C) = 1$. Introducing the partition function

$$Z = \sum_C e^{-E(C)/T}, \quad (1.57)$$

one can then eventually rewrite the distribution $P(C)$ in the form

$$P(C) = \frac{1}{Z} e^{-E(C)/T}, \quad (1.58)$$

which is the standard form of the canonical distribution.

The partition function Z is a useful tool in statistical physics. For instance, the average energy $\langle E \rangle$ can be easily computed from Z :

$$\begin{aligned}
\langle E \rangle &= \sum_C P(C) E(C) = \sum_C E(C) \frac{1}{Z} e^{-E(C)/T} \\
&= \frac{1}{Z} \sum_C E(C) e^{-\beta E(C)} \\
&= -\frac{1}{Z} \frac{\partial Z}{\partial \beta} = -\frac{\partial \ln Z}{\partial \beta}.
\end{aligned} \tag{1.59}$$

Instead of Z , one may also use the “free energy” F defined as

$$F = -T \ln Z. \tag{1.60}$$

Let us give a simple example of computation of Z , in the case of the paramagnetic spin model. The partition function is given by

$$Z = \sum_{\{s_i = \pm 1\}} e^{-\beta E(\{s_i\})}, \tag{1.61}$$

with $E(\{s_i\}) = -h \sum_{i=1}^N s_i$. Hence one has

$$\begin{aligned}
Z &= \sum_{\{s_i = \pm 1\}} e^{\beta h \sum_{i=1}^N s_i} \\
&= \sum_{\{s_i = \pm 1\}} \prod_{i=1}^N e^{\beta h s_i} = \prod_{i=1}^N \left(\sum_{s = \pm 1} e^{\beta h s} \right)
\end{aligned} \tag{1.62}$$

so that one finds

$$Z = (e^{\beta h} + e^{-\beta h})^N. \tag{1.63}$$

Turning to the average energy, one has

$$\langle E \rangle = -\frac{\partial \ln Z}{\partial \beta} = -N \frac{\partial}{\partial \beta} \ln (e^{\beta h} + e^{-\beta h}), \tag{1.64}$$

so that one obtains, recalling that $\beta = 1/T$,

$$\langle E \rangle = -Nh \tanh \frac{h}{T}. \tag{1.65}$$

It is interesting to note that the above equation has exactly the same form as Eq. (1.49), provided that one replaces E , which is fixed in the microcanonical ensemble, by its average value $\langle E \rangle$ in the canonical ensemble. This property is an example of a general property called the “equivalence of ensembles”: in the limit of large systems, the relations between macroscopic quantities are the same in the different statistical ensembles, regardless of which quantity is fixed and which one is fluctuating through

exchanges with a reservoir. The interpretation of this important property is basically that fluctuating observables actually have very small relative fluctuations for large system sizes. This property is deeply related to the Law of Large Numbers and to the Central Limit Theorem—see Chap. 6. Indeed, the relative fluctuations (quantified by the standard deviation normalized by the number of terms) of a sum of independent and identically distributed random variables go to zero when the number of terms in the sum goes to infinity. Note that the equivalence of ensembles generally breaks down in the presence of long-range interactions in the systems.

Another example where the computation of Z is straightforward is the perfect gas. In this case, one has

$$\begin{aligned} Z &= \int_0^L dq_1 \dots \int_0^L dq_{3N} \int_{-\infty}^{\infty} dp_1 \dots \int_{-\infty}^{\infty} dp_{3N} e^{-\beta \sum_{j=1}^{3N} p_j^2/2m} \\ &= L^{3N} \prod_{j=1}^{3N} \int_{-\infty}^{\infty} dp_j e^{-\beta p_j^2/2m} \\ &= V^N \left(\int_{-\infty}^{\infty} dp e^{-\beta p^2/2m} \right)^{3N}. \end{aligned} \quad (1.66)$$

Given that

$$\int_{-\infty}^{\infty} dp e^{-\beta p^2/2m} = \sqrt{\frac{2\pi m}{\beta}}, \quad (1.67)$$

one finds

$$Z = V^N \left(\frac{2\pi m}{\beta} \right)^{\frac{3N}{2}}. \quad (1.68)$$

Computing the average energy leads to

$$\langle E \rangle = -\frac{\partial \ln Z}{\partial \beta} = \frac{3N}{2\beta} = \frac{3}{2}NT \quad (1.69)$$

yielding another example of ensemble equivalence, as this result has the same form as Eq. (1.45). Equation (1.69) is also an example of the general relation of energy equipartition, valid for all quadratic degrees of freedom. More precisely, the equipartition relation states that, in the canonical ensemble, any individual degree of freedom x with a quadratic energy $\frac{1}{2}\lambda x^2$ has an average energy

$$\left\langle \frac{1}{2}\lambda x^2 \right\rangle = \frac{1}{2}k_B T, \quad (1.70)$$

where we have temporarily reintroduced the Boltzmann constant k_B (otherwise set to $k_B = 1$) to comply with standard formulations.

1.3.2 Canonical Entropy

As we have seen above, the microcanonical entropy is defined as $S(E) = \ln \Omega(E)$. This definition is clearly related to the equiprobability of accessible microscopic configurations, since it is based on a counting of accessible configurations. A natural question is then to know how to define the entropy in more general situations. A generic definition of entropy has appeared in information theory, namely:

$$S = - \sum_C P(C) \ln P(C) \quad (1.71)$$

where the sum is over all accessible configurations of the system. This entropy is called the Boltzmann-Gibbs, von Neumann or Shannon entropy depending on the context. This definition of entropy is moreover consistent with the microcanonical one: if $P(C) = 1/\Omega(E)$ for configurations of energy E , and $P(C) = 0$ otherwise, one finds:

$$S = - \sum_{C:E(C)=E} \frac{1}{\Omega(E)} \ln \frac{1}{\Omega(E)} = \ln \Omega(E). \quad (1.72)$$

In this general framework, the canonical entropy reads

$$\begin{aligned} S_{\text{can}} &= - \sum_C P_{\text{can}}(C) \ln P_{\text{can}}(C) \\ &= \sum_C \frac{1}{Z} e^{-\beta E(C)} (\ln Z + \beta E(C)) \\ &= \ln Z + \beta \langle E \rangle. \end{aligned} \quad (1.73)$$

Recalling that the free energy F is defined as $F = -T \ln Z$, one thus has $TS = -F + \langle E \rangle$, which is nothing but the well-known relation $F = \langle E \rangle - TS$. Another standard thermodynamic relation may be found using $\langle E \rangle = -\partial \ln Z / \partial \beta$:

$$\begin{aligned} S &= \ln Z - \beta \frac{\partial \ln Z}{\partial \beta} \\ &= \ln Z + T \frac{\partial \ln Z}{\partial T} \\ &= \frac{\partial}{\partial T} (T \ln Z) \end{aligned} \quad (1.74)$$

so that one finds the standard thermodynamic relation

$$S_{\text{can}} = - \frac{\partial F}{\partial T}. \quad (1.75)$$

1.3.3 Exchanges of Particles with a Reservoir: The Grand-Canonical Ensemble

Similarly to what was done to obtain the canonical ensemble from the microcanonical one by allowing energy exchanges with a reservoir, one can further allow exchanges of particles with a reservoir. The corresponding situation is called the grand-canonical ensemble.

We thus consider a macroscopic system \mathcal{S} exchanging both energy and particles with a reservoir \mathcal{R} . The total system $\mathcal{S} \cup \mathcal{R}$ is isolated with total energy E_{tot} and total number of particles N_{tot} fixed:

$$E + E_{\mathcal{R}} = E_{\text{tot}}, \quad N + N_{\mathcal{R}} = N_{\text{tot}}. \quad (1.76)$$

Generalizing the calculations made in the canonical case, one has (with K a normalization constant),

$$\begin{aligned} P_{\text{GC}}(C) &= K \Omega_{\mathcal{R}}(E_{\mathcal{R}}, N_{\mathcal{R}}) \\ &= K \Omega_{\mathcal{R}}(E_{\text{tot}} - E(C), N_{\text{tot}} - N(C)) \\ &= K \exp[S_{\mathcal{R}}(E_{\text{tot}} - E(C), N_{\text{tot}} - N(C))]. \end{aligned} \quad (1.77)$$

As $E(C) \ll E_{\text{tot}}$ and $N(C) \ll N_{\text{tot}}$, one can expand the entropy $S_{\mathcal{R}}(E_{\text{tot}} - E(C), N_{\text{tot}} - N(C))$ to first order:

$$\begin{aligned} S_{\mathcal{R}}(E_{\text{tot}} - E(C), N_{\text{tot}} - N(C)) &= S_{\mathcal{R}}(E_{\text{tot}}, N_{\text{tot}}) \\ &\quad - E(C) \left. \frac{\partial S_{\mathcal{R}}}{\partial E_{\mathcal{R}}} \right|_{E_{\text{tot}}, N_{\text{tot}}} - N(C) \left. \frac{\partial S_{\mathcal{R}}}{\partial N_{\mathcal{R}}} \right|_{E_{\text{tot}}, N_{\text{tot}}}. \end{aligned} \quad (1.78)$$

As before, the derivative $\partial S_{\mathcal{R}} / \partial E_{\mathcal{R}}$ is identified with $1/T$. We also introduce a new parameter, the chemical potential μ , defined as:

$$\mu = -T \frac{\partial S_{\mathcal{R}}}{\partial N_{\mathcal{R}}} \quad (1.79)$$

(the T factor is conventional). Similarly to the temperature which takes equal values when subsystems exchanging energy have reached equilibrium, the chemical potential takes equal values in subsystems exchanging particles, when equilibrium is attained. Gathering all the above results and notations, one finds that

$$P_{\text{GC}}(C) = \frac{1}{Z_{\text{GC}}} \exp\left(-\frac{1}{T} E(C) + \frac{\mu}{T} N(C)\right) \quad (1.80)$$

which is the standard form of the so-called grand-canonical distribution. The normalization constant Z_{GC} , defined by

$$Z_{\text{GC}} = \sum_C \exp\left(-\frac{1}{T}E(C) + \frac{\mu}{T}N(C)\right), \quad (1.81)$$

is called the grand-canonical partition function.

1.4 Phase Transitions and Ising Model

Phase transitions correspond to a sudden change of behavior of the system when varying an external parameter across a transition point. This could be of interest in complex systems well beyond physics, and is generically associated with collective effects. To illustrate this last property, let us briefly come back to the paramagnetic model defined in Sect. 1.2.3, for which the mean magnetization per spin is given by

$$\langle m \rangle \equiv \frac{\langle M \rangle}{N} = \tanh\left(\frac{h}{T}\right). \quad (1.82)$$

The magnetization is non-zero only if there is a non-zero external field which tends to align the spins. A natural question is thus to know whether one could obtain a non-zero magnetization by including interactions tending to align spins between them (and not with respect to an external source). In this spirit, let us consider the standard (interaction) energy of the Ising model, in the absence of external field:

$$E_{\text{Ising}} = -J \sum_{\langle i,j \rangle} s_i s_j, \quad J > 0. \quad (1.83)$$

This interaction energy is minimized when all spins are parallel. To compute the mean magnetization per spin, one would need to compute either the partition function in presence of an external magnetic field and take the derivative of the free energy with respect to the field, or to compute directly the mean magnetization from its definition. In any case, this is a very complicated task as soon as the space dimension D is larger than one, and the exact calculation has been achieved only in dimensions one and two. The results can be summarized as follows:

- $D = 1$: $m = 0$ for all $T > 0$, so that there is no phase transition at finite temperature. Calculations are relatively easy.
- $D = 2$: there is a phase transition at a finite critical temperature T_c , namely $m = 0$ for $T \geq T_c$ and $m \neq 0$ for $T < T_c$. Calculations are however very technical.
- $D \geq 3$: no analytical solution is known, but numerical simulations show that there is a phase transition at a finite temperature that depends on D .

1.4.1 Ising Model in Fully Connected Geometry

An interesting benchmark model, which can be shown analytically to exhibit a phase transition, is the fully connected Ising model, whose energy is defined as

$$E_{\text{fc}} = -\frac{J}{N} \sum_{i<j} s_i s_j + E_0, \quad (1.84)$$

where the sum is over all pairs of spins in the system. The $1/N$ prefactor is included in order to keep the energy per spin finite in the large N limit. The term E_0 is added for later convenience, and is arbitrary at this stage (it does not modify the canonical distribution). Considering the magnetization $M = \sum_{i=1}^N s_i$, one has, given that $s_i^2 = 1$,

$$M^2 = 2 \sum_{i<j} s_i s_j + N \quad (1.85)$$

from which one finds

$$E_{\text{fc}} = -\frac{J}{2N}(M^2 - N) + E_0 = -\frac{J}{2N}M^2 + \frac{J}{2} + E_0. \quad (1.86)$$

Choosing $E_0 = -J/2$, and introducing the magnetization per spin $m = M/N$, one finds

$$E_{\text{fc}} = -\frac{J}{2}Nm^2. \quad (1.87)$$

One possible way to detect the phase transition is to compute the probability distribution $P(m)$ of the magnetization, by summing over all configurations having a given magnetization m :

$$\begin{aligned} P(m) &= \frac{1}{Z} \sum_{C:m(C)=m} e^{-\beta E(C)} \\ &= \frac{1}{Z} e^{S(m) + \frac{1}{2}\beta JNm^2} \end{aligned} \quad (1.88)$$

where $\Omega(m) = e^{S(m)}$ is the number of configurations with magnetization m . Using the relation

$$\Omega(m) = \frac{N!}{N_+!N_-!} \quad (1.89)$$

with

$$N_+ = \frac{N}{2}(1+m), \quad N_- = \frac{N}{2}(1-m), \quad (1.90)$$

one obtains for $S(m) = \ln \Omega(m)$

$$S(m) = -N \left[\frac{1+m}{2} \ln(1+m) + \frac{1-m}{2} \ln(1-m) - \ln 2 \right]. \quad (1.91)$$

Hence from Eqs. (1.88) and (1.91) it turns out that $P(m)$ can be written as

$$P(m) = e^{-Nf(m)} \quad (1.92)$$

with $f(m)$ given by

$$f(m) = \frac{1+m}{2} \ln(1+m) + \frac{1-m}{2} \ln(1-m) - \frac{J}{2T} m^2 + f_0(T), \quad (1.93)$$

where $f_0(T)$ is a temperature-dependent constant, ensuring that the minimum value reached by $f(m)$ is 0, to be consistent with the normalization of $P(m)$. The function $f(m)$ is called a large deviation function, or a Landau free energy function. Hence the magnetization m_0 that maximizes the probability distribution $P(m)$ corresponds to a minimum of $f(m)$. Moreover, fluctuations around m_0 are exponentially suppressed with N . For high temperature T , the term J/T is small, and the entropic contribution to $f(m)$ should dominate, leading to $m_0 = 0$. To understand what happens when temperature is progressively lowered, it is useful to expand $f(m)$ for small values of m , up to order m^4 , leading to:

$$f(m) = f_0(T) + \frac{1}{2} \left(1 - \frac{J}{T} \right) m^2 + \frac{1}{12} m^4 + \mathcal{O}(m^6). \quad (1.94)$$

One can then distinguish two different cases:

- If $T \geq T_c \equiv J$, $f(m)$ has only one minimum, for $m = 0$.
- If $T < T_c$, $f(m)$ has two symmetric minima $\pm m_0$. These minima are obtained as solutions of the equation $df/dm = 0$:

$$\frac{df}{dm} = \left(1 - \frac{J}{T} \right) m + \frac{1}{3} m^3 = - \left| 1 - \frac{J}{T} \right| m + \frac{1}{3} m^3 = 0. \quad (1.95)$$

The non-zero solutions are $m = \pm m_0$ with

$$m_0 = \sqrt{3 \left(\frac{J}{T} - 1 \right)} = \sqrt{3} \left(\frac{T_c - T}{T} \right)^{1/2}. \quad (1.96)$$

It can be checked easily that the solution $m = 0$ corresponds in this case to a local maximum of $f(m)$, and thus to a local minimum of $P(m)$ (Fig. 1.1).

Hence, there is a phase transition at $T = T_c \equiv J$, T_c being called the critical temperature. The most probable magnetization m_0 is called the “order parameter of the phase transition”, as the phase transition is characterized by the onset of a non-zero value of m_0 . In addition, the order parameter varies as $m_0 \sim (T_c - T)^\beta$

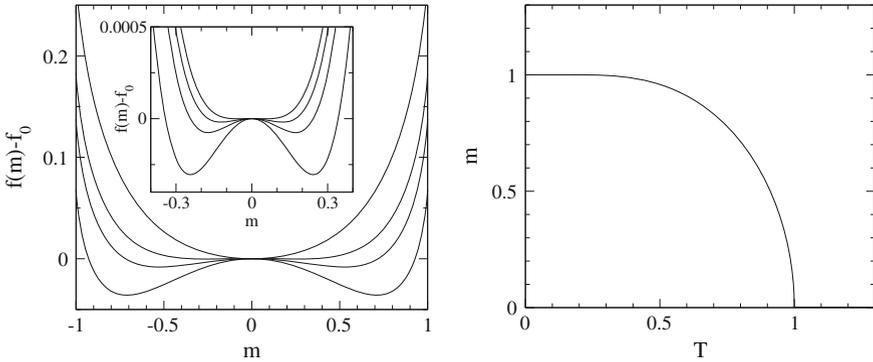


Fig. 1.1 *Left, main plot* large deviation function $f(m)$, for temperature $T = 1.2, 0.98, 0.9$, and 0.8 from *top to bottom* ($T_c = 1$). Two symmetric minima appear for $T < T_c$, indicating the onset of magnetized states. *Inset* zoom on the temperature range close to T_c ; $f(m)$ is plotted for $T = 0.999, 0.995, 0.99$ and 0.98 from *top to bottom*. *Right* magnetization $m(T)$ as a function of temperature ($T_c = 1$)

for T close to T_c ($T < T_c$), with $\beta = 1/2$ here. The exponent β is an example of critical exponent, and the value $\beta = 1/2$ is called the “mean-field value” of β , for reasons that will become clear in the next section. The notation β is standard for the critical exponent associated to the order parameter, and should not be confused with the inverse temperature $\beta = 1/T$.

An important remark is that the average value $\langle m \rangle$ of the magnetization is still zero for $T < T_c$, since the two values $\pm m_0$ of the magnetization have the same probability. However, for a large system, the time needed to switch between states m_0 and $-m_0$ becomes very large (at least if one uses a local spin-flip dynamics), and the time-averaged magnetization over a typical observation time window is non-zero, and equal either to m_0 or to $-m_0$.

1.4.2 Ising Model with Finite Connectivity

We now come back to the Ising model in a finite-dimensional space of dimension D . As mentioned above, the analytical solution is hard to obtain in dimension $D = 2$, and is not known in higher dimensions. However, useful approximations have been developed, the most famous one being called the mean-field approximation.

The reason why the fully connected model can be easily solved analytically is that its energy E is a function of the magnetization m only, as seen in Eq. (1.87). When the model is defined on a finite-dimensional lattice, this property is no longer true, and the energy reads:

$$E = -\frac{J}{2} \sum_{i=1}^N s_i \left(\sum_{j \in \mathcal{V}(i)} s_j \right) \quad (1.97)$$

where $\mathcal{V}(i)$ is the set of neighboring sites of site i . The factor $1/2$ comes from the fact that a given link of the lattice now appears twice in the sum. This last expression can be rewritten as

$$E = -DJ \sum_{i=1}^N s_i \langle s_j \rangle_{\mathcal{V}(i)}, \quad (1.98)$$

$\langle s_j \rangle_{\mathcal{V}(i)}$ being the local magnetization per spin of the set of neighbors $\mathcal{V}(i)$:

$$\langle s_j \rangle_{\mathcal{V}(i)} \equiv \frac{1}{2D} \sum_{j \in \mathcal{V}(i)} s_j. \quad (1.99)$$

The parameter D is the space dimension, and the number of neighbors of a given site i is $2D$, given that we consider hypercubic lattices (square lattice in $D = 2$, cubic lattice in $D = 3, \dots$).

As a first approximation, one could replace the local magnetization per spin of the set of neighbors by the global magnetization per spin of the whole system, $m = N^{-1} \sum_{i=1}^N s_i$:

$$\langle s_j \rangle_{\mathcal{V}(i)} \rightarrow m. \quad (1.100)$$

This approximation leads to the following expression of the energy

$$E \approx E_{\text{mf}} = -DJm \sum_{i=1}^N s_i = -DJNm^2, \quad (1.101)$$

where the subscript ‘mf’ stands for “mean-field” approximation. Then E_{mf} depends only on the magnetization m , and has a form similar to the energy E_{fc} of the fully connected model. One can define an effective coupling $J_{\text{mf}} = 2DJ$ so that the forms of the two energies become exactly the same, namely

$$E_{\text{mf}} = -\frac{1}{2} J_{\text{mf}} Nm^2. \quad (1.102)$$

Now it is clear that the results of the fully connected model can be applied to the present mean-field approximation, yielding a phase transition at $T_c^{\text{mf}} = J_{\text{mf}} = 2DJ$. For $T > T_c^{\text{mf}}$, $\langle m \rangle = 0$ while for $T < T_c^{\text{mf}}$, but close to T_c^{mf} , $\langle m \rangle \sim (T_c^{\text{mf}} - T)^{1/2}$. Qualitatively, the approximation is expected to be valid for large space dimension D . It can be shown, using more involved arguments, that for $D \geq 4$, the approximation is semi-quantitatively valid, in the sense that the value $\beta = 1/2$ of the critical exponent, obtained from the approximation, is correct. However, the value of the critical temperature T_c^{mf} is not correctly predicted by the mean-field approximation,

namely $T_c \neq T_c^{\text{mf}}$. For $D < 4$, the value of β differs from the mean-field value $1/2$, and the mean-field approximation breaks down. For $D = 3$, numerical simulations indicate that $\beta \approx 0.31$, and for $D = 2$, the exact solution yields $\beta = 1/8$. Finally, for $D = 1$, $\langle m \rangle = 0$ except for $T = 0$, so that the exponent β is not defined [6].

The discrepancy mentioned above between mean-field predictions and results obtained in low-dimensional systems mainly comes from the presence of fluctuations of the local magnetization $\sum_{j \in \mathcal{V}(i)} s_j$. Since on the other hand exact solutions are very hard to obtain, there is need for a different approach, that could be generic enough and could be centered on the issue of correlation, which is at the heart of the difficulties encountered. This is precisely the aim of the renormalization group approach.

1.4.3 Renormalization Group Approach

A standard observation on finite dimensional systems exhibiting a continuous phase transition is that the correlation length diverges when the temperature approaches the critical temperature T_c . The correlation length is defined through the correlation function

$$C_{ij} = \langle (s_i - m_0)(s_j - m_0) \rangle = \langle s_i s_j \rangle - m_0^2. \quad (1.103)$$

As soon as the distance $r = d_{ij}$ between sites i and j is large with respect to the lattice spacing a , the correlation function generally becomes isotropic, $C_{ij} = C(r)$. In addition, the large distance behavior of $C(r)$ is often of the form

$$C(r) \sim \frac{1}{r^\alpha} e^{-r/\xi}, \quad \alpha > 0, \quad (1.104)$$

which defines the correlation length ξ . The latter diverges for $T \rightarrow T_c$. This is the reason why direct calculations in the range $T \approx T_c$ are very difficult, due to the strong correlation between spins. A natural idea is to look for an approach that could reduce in some way the intensity of correlations, in order to make calculations tractable.

This is basically the principle of the renormalization group (RG) approach, in which one progressively integrates out small scale degrees of freedom. The idea is that at the critical point, structures are present at all scales, from the lattice spacing to the system size. A RG transform may intuitively be thought of as defocusing the picture of the system, so that fine details become blurred. This method is actually very general, and could be relevant in many fields of complex system sciences, given that issues like large scale correlations and scale invariance or fractals are often involved in complex systems.

For definiteness, let us however consider again the Ising model. To implement the RG ideas in a practical way, one could make blocks of spins and define an effective spin for each block, with effective interactions with the neighboring blocks. The effective interactions are defined in such a way that the large scale properties

are the same as for the original (non-renormalized) model. This is done in practice by conserving the partition function, namely $Z' = Z$ (in the present section, the prime denotes renormalized quantities). One would then like to define a renormalized interaction constant J' such that

$$H' = -J' \sum_{\langle B_1, B_2 \rangle} S_{B_1} S_{B_2} \quad (1.105)$$

where B_1 and B_2 are generic labels for the blocks (the sites of the renormalized lattice). The problem is that very often, the RG transform generates new effective couplings, like next-nearest-neighbor couplings, that were absent in the original model, and the number of couplings keeps increasing with the number of iterations of the RG transform. However, in some simple cases, the transformation can be performed exactly, without increasing the number of coupling constants, as we shall see later on.

Yet, let us first emphasize the practical interest of the RG transform. We already mentioned that one of the main difficulties comes from the presence of long-range correlations close to the critical point. Through the RG transform, the lattice spacing becomes $a' = 2a$ (if one makes blocks of linear size $2a$). On the contrary, the correlation length remains unchanged, since the large scale properties remain unaffected by the RG transform. Hence the correlation length expressed in unit of the lattice spacing, namely ξ/a , decreases by a factor of 2 in the transformation, to become

$$\frac{\xi'}{a'} = \frac{1}{2} \frac{\xi}{a}. \quad (1.106)$$

Thus upon iterations of the RG transform, the effective Hamiltonian becomes such that $\xi' \sim a'$, so that standard approximation schemes (mean-field,...) can be used. One then needs to follow the evolution of the coupling constant J' under iterations. This is called the renormalization flow.

An explicit example can be given with the one-dimensional Ising chain, using a specific RG transform called decimation procedure [7]. We start with the energy (or Hamiltonian)

$$H = \sum_{i=1}^N H_{i,i+1}(s_i, s_{i+1}) \quad (1.107)$$

where the local interaction term $H_{i,i+1}(s_i, s_{i+1})$ is given by

$$H_{i,i+1}(s_i, s_{i+1}) = -J s_i s_{i+1} + c. \quad (1.108)$$

Note that periodic boundary conditions are understood. The constant c plays no role at this stage, but it will be useful later on in the renormalization procedure. The basic idea of the decimation procedure is to perform, in the partition function, a partial sum over the spins of—say—odd indices in order to define renormalized coupling

constants J' and h' . Then summing over the values of the spins with even indices yields the partition function Z' of the renormalized model, which is by definition of the renormalization procedure equal to the initial partition function Z . To be more explicit, one can write Z as

$$Z = \sum_{\{s_{2j}\}} \sum_{\{s_{2j+1}\}} \exp[-\beta H(\{s_i\})] \quad (1.109)$$

where $\sum_{\{s_{2j}\}}$ (resp. $\sum_{\{s_{2j+1}\}}$) indicates a sum over all possible values of the $N/2$ variables $\{s_{2j}\}$ (resp. $\{s_{2j+1}\}$). Equation (1.109) can then be rewritten in the following form:

$$Z = \sum_{\{s_{2j}\}} \exp[-\beta H'(\{s_{2j}\})] \quad (1.110)$$

where $H'(\{s_{2j}\})$ is the renormalized Hamiltonian, defined by

$$\exp[-\beta H'(\{s_{2j}\})] = \sum_{\{s_{2j+1}\}} \exp[-\beta H(\{s_i\})]. \quad (1.111)$$

Assuming that the renormalized Hamiltonian can be decomposed into a sum of local terms

$$H'(\{s_{2j}\}) = \sum_{j=1}^{N/2} H'_{j,j+1}(s_{2j}, s_{2j+2}) \quad (1.112)$$

we get from Eq. (1.111) the relation

$$\begin{aligned} & \prod_{j=1}^{N/2} \exp[-\beta H'_{j,j+1}(s_{2j}, s_{2j+2})] \\ &= \sum_{\{s_{2j+1}\}} \prod_{j=1}^{N/2} \exp[-\beta H_{2j,2j+1}(s_{2j}, s_{2j+1}) - \beta H_{2j+1,2j+2}(s_{2j+1}, s_{2j+2})] \\ &= \prod_{j=1}^{N/2} \sum_{s_{2j+1}} \exp[-\beta H_{2j,2j+1}(s_{2j}, s_{2j+1}) - \beta H_{2j+1,2j+2}(s_{2j+1}, s_{2j+2})] \end{aligned} \quad (1.113)$$

where in the last line, the sum runs over the single variable s_{2j+1} , the index j being fixed within the product. This last relation is satisfied if, for any given $j = 1, \dots, N/2$, and any given values of s_{2j} and s_{2j+2} ,

$$\begin{aligned} & \exp[-\beta H'_{j,j+1}(s_{2j}, s_{2j+2})] \\ &= \sum_{s_{2j+1}=\pm 1} \exp[-\beta H_{2j,2j+1}(s_{2j}, s_{2j+1}) - \beta H_{2j+1,2j+2}(s_{2j+1}, s_{2j+2})]. \end{aligned} \quad (1.114)$$

Further assuming that $H'_{j,j+1}(s_{2j}, s_{2j+2})$ takes the form

$$H'_{j,j+1}(s_{2j}, s_{2j+2}) = -J' s_{2j} s_{2j+2} + c', \quad (1.115)$$

where J' and c' are the renormalized parameters, one obtains

$$\exp[\beta J' s_{2j} s_{2j+2} - \beta c'] = \sum_{s_{2j+1}=\pm 1} \exp[\beta J(s_{2j} s_{2j+1} + s_{2j+1} s_{2j+2}) - 2\beta c]. \quad (1.116)$$

Introducing the reduced variable³

$$u = e^{-4\beta J}, \quad (1.117)$$

Equation (1.116) leads to the following recursion relation:

$$u' = \frac{4u}{(1+u)^2}. \quad (1.118)$$

Let us denote as ξ_{nd} the dimensionless correlation length

$$\xi_{\text{nd}} = \frac{\xi}{a}. \quad (1.119)$$

Then from Eq. (1.106) the recursion relation for ξ_{nd} reads

$$\xi'_{\text{nd}} = \frac{1}{2} \xi_{\text{nd}}, \quad (1.120)$$

from which one deduces that the fixed points of the renormalization procedure, that satisfy $\xi'_{\text{nd}} = \xi_{\text{nd}}$, can only be $\xi_{\text{nd}} = \infty$ or $\xi_{\text{nd}} = 0$. The latter is called the trivial fixed point, as it corresponds to the limit situation where no correlation is present in the system. In contrast, the fixed point $\xi_{\text{nd}} = \infty$ corresponds to the critical fixed point, where correlation extends over the whole system size. As ξ_{nd} decreases through iteration of the RG transform, the critical fixed point $\xi_{\text{nd}} = \infty$ is unstable, while the trivial fixed point $\xi_{\text{nd}} = 0$ is stable.

Coming back to the iteration relation Eq. (1.118), let us first look for the fixed points of this equation, namely the solutions of

$$u = \frac{4u}{(1+u)^2}. \quad (1.121)$$

The value $u = 0$ is obviously a solution, and it is easy to check that $u = 1$ is the other positive solution ($u = -3$ is the third solution, but in view of Eq. (1.117), we

³We do not follow here the evolution of the constant c under renormalization, and rather focus on the evolution of the physically relevant coupling constant J .

are seeking for positive solutions only). Then to identify which one of the two fixed points is the critical point, we need to investigate the stability of each fixed point under iteration. The stability is studied by introducing a small variation δu around a given fixed point u_1 , namely $u = u_1 \pm \delta u$, and writing the evolution equation for δu to leading order. For $u_1 = 0$, one finds, with $u = \delta u$,

$$\delta u' = \frac{4\delta u}{(1 + \delta u)^2} \approx 4\delta u, \quad \delta u > 0, \quad (1.122)$$

so that δu increases upon iteration: the fixed point $u_1 = 0$ is unstable, and thus corresponds to the critical fixed point. Besides, the fixed point $u_1 = 1$ is easily checked to be stable. Using $u = 1 - \delta u$, we have

$$1 - \delta u' = \frac{4(1 - \delta u)}{(2 - \delta u)^2}, \quad (1.123)$$

leading after a second order expansion in δu to

$$\delta u' \approx \frac{1}{4}\delta u^2. \quad (1.124)$$

Hence δu converges to 0 upon iteration, confirming the stability of the fixed point $u_1 = 1$. Coming back to the critical fixed point, and recalling the definition Eq. (1.117), one sees that $u_1 = 0$ corresponds to an infinite value of J/T . In the above framework, this case is interpreted as an infinite coupling limit, as the iteration was made on J . However, the fixed point can also be interpreted as a zero-temperature fixed point, keeping the coupling constant J fixed. A sketch of the corresponding renormalization flow is presented in the top panel of Fig. 1.2.

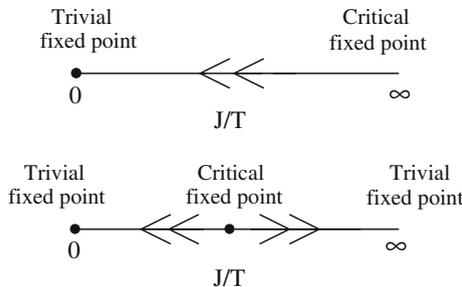


Fig. 1.2 Sketch of the renormalization flow, in terms of the reduced coupling constant J/T . In all cases, the zero coupling (or infinite temperature) point is a trivial fixed point, but the position of the critical fixed point may differ from one case to the other. *Top* one-dimensional Ising model; the critical fixed point corresponds to infinite coupling (or zero temperature). *Bottom* fully connected Ising model, or Ising model in dimension $D \geq 2$; the critical fixed point corresponds to a finite value of the reduced coupling, implying a finite critical temperature for a given coupling

This one-dimensional example is of course only a very simple case, which can be solved through other more direct methods. However, it is a good illustration of the way the concept of RG can be implemented in practice. In two- or three-dimensional models, exact treatments like the above one are most often not available. Yet, many approaches based on different approximation schemes have been developed. A typical situation in dimension $D > 1$ is that there is a finite value K_c of the ratio $K = J/T$ which corresponds to a critical fixed point, and both values $K = 0$ and $K = \infty$ correspond to trivial fixed points, where no correlation is present (see bottom panel of Fig. 1.2). Quite importantly, linearizing the iteration equation in the vicinity of the critical fixed point allows the determination of the critical exponent β , as well as other critical exponents. In the Ising chain studied above, this is not possible because the critical temperature is zero, so that there is no extended temperature region where the magnetization is non-zero. But this approach turns out to be relevant in dimension higher than one.

As an illustration of the emergence of a critical fixed point with a finite coupling K_c , let us briefly consider again the fully-connected Ising model (which, as seen above, can be studied by more direct means than the renormalization group method). The energy of the fully-connected Ising model reads $H_{fc} = -\frac{1}{2}JNm^2 + c$, where c is an arbitrary constant. Below, we denote as $K = \beta J$ the dimensionless coupling constant (with β the inverse temperature). We perform a very simple renormalization procedure that consists in integrating out a single spin, going from a system of $N + 1$ spins to a system of N spins. Denoting as \tilde{K} and \tilde{c} the parameters of the renormalized system, the renormalization transformation reads

$$\exp\left(\frac{1}{2}\tilde{K}Nm^2 + \tilde{c}\right) = \sum_{S_{N+1}=\pm 1} \exp\left(\frac{1}{2}\frac{K}{N+1}(Nm + S_{N+1})^2 + c\right) \quad (1.125)$$

with $m = N^{-1} \sum_{i=1}^N S_i$. Expanding the square in the r.h.s. of Eq. (1.125) and taking the logarithm of both sides, one obtains in the large N limit

$$\frac{1}{2}(\tilde{K} - K)Nm^2 + \tilde{c} = -\frac{1}{2}Km^2 + \ln \cosh(Km) + \ln 2 + c. \quad (1.126)$$

As for the one-dimensional Ising model, we do not follow the constant term c and focus on the coupling constant K . Expanding the hyperbolic cosine to order m^2 , we get from the coefficients of the m^2 terms

$$(\tilde{K} - K)N = -K + K^2. \quad (1.127)$$

Note that higher order terms in m , like m^4 , are also generated in this transformation; however, we do not study their role here. The renormalization flow is often characterized by a parameter ℓ which is additive when several renormalization transformations are successively performed. In a single transformation, this parameter varies by $\delta\ell = \ln b$, where $b > 1$ is the scale factor of the transformation. Here,

$b = (N + 1)/N$, so that for large N , $\delta\ell = 1/N$. Considering K as a function of ℓ , we can thus rewrite Eq. (1.127) as

$$\frac{dK}{d\ell} = -K + K^2. \quad (1.128)$$

It is easy to check that $K_c = 1$ is a critical (unstable) fixed point, while $K = 0$ and $K = \infty$ are trivial (stable) fixed points. Hence we recover the phase transition at $K_c = 1$ obtained by direct calculations in Sect. 1.4.1.

1.5 Disordered Systems and Glass Transition

In the general framework of complex systems, disordered systems are systems where each particle or agent has specific properties, which are qualitatively the same for all of them, but differ quantitatively from one to the other. In theoretical models, these quantitative properties are most often drawn from a given probability distribution for each particle or agent, and remain constant in time. Disordered systems should be very relevant in complex systems studies, like social science for instance, as each human being has its specific skills or tastes.

In a physical context, the concept of disordered system could seem less natural, since all particles within a given class are identical. In this case, disorder rather comes from the possibly random position of the particles. A standard example is that of magnetic impurities (that carry a magnetic moment, or spin) diluted in a non-magnetic material. The interaction between magnetic atoms (which have to be described in the framework of quantum mechanics) is mediated by the non-magnetic atoms, and acquires an oscillatory behavior, depending on the distance r_{ij} between the two spins:

$$H = - \sum_{i,j} J(r_{ij}) s_i s_j. \quad (1.129)$$

The interaction constant $J(r_{ij})$ is a given function of the distance r_{ij} , which oscillates around 0, thus taking both positive and negative values. The amplitude of the oscillations decays as a power-law of the distance. As the distances between atoms are random, the interactions between atoms have a random sign, which is the basic property of spin-glasses.

1.5.1 Theoretical Spin-Glass Models

In order to propose a simplified model for spin-glass materials, it has been proposed to replace the positional disorder by an interaction disorder, the magnetic atoms being now situated on a regular lattice. To this purpose, one considers an Ising-like

model in D -dimensions, where the spins are placed at each node of a lattice. Spins on neighboring sites (i, j) interact through a coupling constant J_{ij} , drawn from a distribution $P(J)$. As the couplings J_{ij} are kept constant in time, one speaks about quenched disorder. This model is called the Edwards-Anderson model [8]. In $D = 1$, the Edwards-Anderson model is qualitatively equivalent to the standard Ising model, up to a redefinition of the spins. In $D > 1$, analytical solutions are not known, and results have thus been obtained through numerical simulations. A fully connected version, called the Sherrington-Kirkpatrick model, has been proposed and solved [9], but the techniques involved are already rather difficult, even at the level of the fully connected model. The main qualitative picture [10] emerging from these models is that below a given energy level, the phase space decomposes into a lot of valleys, or metastable states, from which it takes a very long time to escape.

In order to give a flavor of the basic properties of spin-glass models, we present below two of the simplest models of spin glasses, namely the Mattis model and the Random Energy Model.

1.5.2 A Toy Model for Spin-Glasses: The Mattis Model

Spin-glass models are very complicated to study analytically, due to the presence of disorder. There is a simplified case, however, in which explicit calculations can be done relatively easily, while preserving some of the main features of spin-glasses. This is the so-called Mattis model [11].

This model consists in a spin-glass on an arbitrary lattice, composed of N spins. The energy of a configuration (s_1, \dots, s_N) takes the usual form

$$H = - \sum_{\langle i,j \rangle} J_{ij} s_i s_j - h \sum_i s_i \quad (1.130)$$

where as usual, the sum is carried over the nearest neighbors on the lattice. The simplification with respect to generic spin glasses comes from the form assumed for the disordered couplings J_{ij}

$$J_{ij} = J \varepsilon_i \varepsilon_j \quad (1.131)$$

where $\varepsilon_i = \pm 1$ are independent quenched random variables. Introducing new spin variables $\sigma_i = \varepsilon_i s_i$, the energy reads

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \varepsilon_i \sigma_i. \quad (1.132)$$

In the following, we shall focus on the case of zero external field, $h = 0$. Let us introduce the mean magnetization per spin m , defined as

$$m \equiv \frac{1}{N} \sum_i \langle s_i \rangle \quad (1.133)$$

where $\langle s_i \rangle$ means a “thermal” average, that is an average over all spin configurations, for given values of the disorder. Then it is easy to show, by averaging over the disorder represented by the variables ε_i , that the disorder-averaged magnetization is equal to zero, $\bar{m} = 0$ (the overbar denotes an average over the disorder).

This can be done as follows. Introducing the variables σ_i , let us rewrite the averaged magnetization as

$$m = \frac{1}{N} \sum_i \varepsilon_i \langle \sigma_i \rangle. \quad (1.134)$$

Performing the average over the disorder, we get

$$\bar{m} = \frac{1}{N} \sum_i \overline{\varepsilon_i \langle \sigma_i \rangle} \quad (1.135)$$

For the case $h = 0$ considered here, the energy expressed in terms of the variables σ takes the standard Ising form

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j \quad (1.136)$$

and is thus independent of the disorder. Hence the thermal average $\langle \sigma_i \rangle$ is also independent of the disorder, so the disorder-averaged magnetization can be rewritten as

$$\bar{m} = \frac{1}{N} \sum_i \overline{\varepsilon_i} \langle \sigma_i \rangle. \quad (1.137)$$

Since $\overline{\varepsilon_i} = 0$, one readily obtains that the disorder-averaged magnetization is equal to zero. Note that this is true whatever the value of $\langle \sigma_i \rangle$; in the low temperature phase $T < T_c$ (where T_c is the critical temperature of the Ising model), $\langle \sigma_i \rangle$ is nonzero, but the resulting magnetization still vanishes once averaged over the disorder.

To go one step further, it is also possible to show using similar methods that the (zero-field) disorder-averaged susceptibility satisfies

$$\chi \equiv \left. \frac{\partial \bar{m}}{\partial h} \right|_{h=0} = \frac{1-q}{k_B T} \quad (1.138)$$

where $q \equiv \overline{\langle s_i \rangle^2} = \langle \sigma_i \rangle^2$. We thus find that the susceptibility diverges at $T = 0$ only, confirming that there is no phase transition at $T > 0$. The parameter q is generically called the Edwards-Anderson order parameter, and it is nonzero as soon as individual spins have nonzero average values, even if these values takes random signs from one spin to the other, and sum up to a zero global magnetization.

1.5.3 The Random Energy Model

Another very simple disordered model, which already captures a lot of the phenomenology of realistic disordered systems, is the Random Energy Model (REM) [12]. The model has 2^N configurations, labeled by an index $\alpha = 1, \dots, 2^N$ (it can be thought of as a spin model, with N spins $s_i = \pm 1$, although space is not explicitly described in this model). To each configuration α is attached a time-independent energy E_α , chosen at random from the distribution

$$P(E) = \frac{1}{\sqrt{N\pi J^2}} \exp\left(-\frac{E^2}{NJ^2}\right). \quad (1.139)$$

All the energies E_α are statistically independent random variables. The variance of E in the distribution (1.139) is taken to be proportional to N in order to ensure that the average energy of the model is extensive, i.e., proportional to N —see below. We denote as $n(E)dE$ the number of configurations with energy in the interval $[E, E + dE]$, so that $n(E)$ is the density of configurations with energy E . The density $n(E)$ is a random quantity, but its fluctuations are small if $\langle n(E) \rangle$ is large, so that $n(E) \approx \langle n(E) \rangle$ in this regime. By definition, $P(E) = \langle n(E) \rangle / 2^N$, so that $\langle n(E) \rangle = 2^N P(E)$, leading to

$$\begin{aligned} \langle n(E) \rangle &= \exp\left(N \ln 2 - \frac{E^2}{NJ^2}\right) \\ &= \exp\left[N \left(\ln 2 - \frac{\varepsilon^2}{J^2}\right)\right] \end{aligned} \quad (1.140)$$

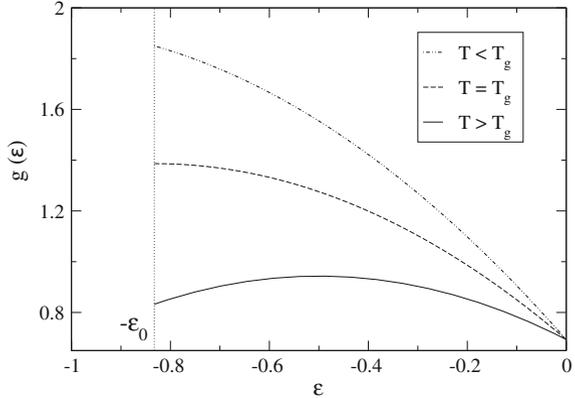
where the energy density $\varepsilon = E/N$ has been introduced. One sees that if $\ln 2 - \varepsilon^2/J^2 > 0$, corresponding to $|\varepsilon| < \varepsilon_0 = J\sqrt{\ln 2}$, $\langle n(E) \rangle$ is exponentially large with N , so that there is a large number of configurations at energy density ε , and the assumption $n(E) \approx \langle n(E) \rangle$ is justified. In contrast, if $\ln 2 - \varepsilon^2/J^2 < 0$, which corresponds to $|\varepsilon| > \varepsilon_0$, $\langle n(E) \rangle$ is exponentially small with N . This means that in most samples, there are no configurations at energy density $|\varepsilon| > \varepsilon_0$. The non-zero, but small value of $\langle n(E) \rangle$ comes from the contribution to the average value of very rare and atypical samples, which include some configurations with exceptionally low (or high) energy.

We can now evaluate the partition function of the REM, defined as

$$Z = \sum_{\alpha=1}^{2^N} e^{-E_\alpha/T}. \quad (1.141)$$

As all the energies E_α are random variables, the partition function Z is also a random variable, which fluctuates from one realization of the disorder to another. Yet, we can evaluate the typical value of Z as follows:

Fig. 1.3 Illustration of the function $g(\varepsilon)$ for $T > T_g$ (full line), $T = T_g$ (dashed line) and $T < T_g$ (dot-dashed), with $J = 1$. The function $g(\varepsilon)$ is plotted over the interval $[-\varepsilon_0, 0]$; the value $\varepsilon = -\varepsilon_0$ is indicated by the vertical dotted line. For $T > T_g$, the maximum ε^* of $g(\varepsilon)$ satisfies $-\varepsilon_0 < \varepsilon^* < 0$, while for $T \leq T_g$, the maximum of $g(\varepsilon)$ over the interval $[-\varepsilon_0, 0]$ is reached at the lower bound $\varepsilon = -\varepsilon_0$



$$Z \approx Z_{\text{typ}} = \int_{-\varepsilon_0}^{\varepsilon_0} d\varepsilon \langle \tilde{n}(\varepsilon) \rangle e^{-N\varepsilon/T}, \quad (1.142)$$

with the notation $\tilde{n}(\varepsilon) = Nn(N\varepsilon)$. In Eq. (1.142), we have replaced $\tilde{n}(\varepsilon)$ by $\langle \tilde{n}(\varepsilon) \rangle$ for $|\varepsilon| < \varepsilon_0$, and by 0 for $|\varepsilon| > \varepsilon_0$, consistently with the above discussion. We can then write, using Eqs. (1.140) and (1.142),

$$Z_{\text{typ}} = \int_{-\varepsilon_0}^{\varepsilon_0} d\varepsilon e^{Ng(\varepsilon)} \quad (1.143)$$

with

$$g(\varepsilon) = \ln 2 - \frac{\varepsilon^2}{J^2} - \frac{\varepsilon}{T}. \quad (1.144)$$

The function $g(\varepsilon)$ is illustrated in Fig. 1.3. In the large N limit, we can evaluate Z_{typ} through a saddle point calculation, namely

$$Z_{\text{typ}} \sim e^{Ng_{\text{max}}(\varepsilon_0)} \quad (1.145)$$

where $g_{\text{max}}(\varepsilon_0)$ is the maximum value of $g(\varepsilon)$ over the interval $[-\varepsilon_0, \varepsilon_0]$ (in practice, over the interval $[-\varepsilon_0, 0]$ since the maximum is always reached for a negative energy value). Let us first consider the maximum ε^* of $g(\varepsilon)$ over the entire real line. Taking the derivative of $g(\varepsilon)$, one has

$$g'(\varepsilon) = -\frac{2\varepsilon}{J^2} - \frac{1}{T}. \quad (1.146)$$

From $g'(\varepsilon) = 0$, we find

$$\varepsilon^* = -\frac{J^2}{2T}. \quad (1.147)$$

As $g(\varepsilon)$ is a parabola with negative curvature, it is increasing for $\varepsilon < \varepsilon^*$ and decreasing for $\varepsilon > \varepsilon^*$. If $\varepsilon^* > -\varepsilon_0$, then $g_{\max}(\varepsilon_0) = g(\varepsilon^*)$, so that

$$Z_{\text{typ}} \sim e^{Ng(\varepsilon^*)}. \quad (1.148)$$

The condition $\varepsilon^* > -\varepsilon_0$ translates into $T > T_g$, where the so-called glass transition temperature T_g is defined as

$$T_g = \frac{J}{2\sqrt{\ln 2}}. \quad (1.149)$$

For $\varepsilon^* < -\varepsilon_0$, or equivalently $T < T_g$, $g(\varepsilon)$ is a decreasing function of ε over the entire interval $[-\varepsilon_0, \varepsilon_0]$, so that $g_{\max}(\varepsilon_0) = g(-\varepsilon_0)$, and

$$Z_{\text{typ}} \sim e^{Ng(-\varepsilon_0)}. \quad (1.150)$$

From these estimates of Z_{typ} , the free energy $F = -T \ln Z_{\text{typ}}$ and the entropy $S = -\partial F / \partial T$ can be computed. For $T > T_g$, one finds

$$F = -N \left(T \ln 2 + \frac{J^2}{4T} \right), \quad (1.151)$$

leading for the entropy to

$$S = N \left(\ln 2 - \frac{J^2}{4T^2} \right). \quad (1.152)$$

For $T < T_g$, we have

$$F = -TNg(-\varepsilon_0) = -NJ\sqrt{\ln 2}. \quad (1.153)$$

The free energy does not depend on temperature in this range, so that the corresponding entropy vanishes:

$$S = -\frac{\partial F}{\partial T} = 0, \quad T < T_g. \quad (1.154)$$

It can also be checked that the entropy given in Eq. (1.152) for $T > T_g$ vanishes continuously for $T \rightarrow T_g$. Hence the temperature T_g corresponds to a glass transition temperature, where the entropy goes to zero when lowering temperature down to T_g , and remains zero below T_g . Actually, to make the statement sharper, only the entropy density S/N goes to zero for $T < T_g$, in the infinite N limit. Computing subleading corrections to the entropy, one finds that the entropy S is independent of N , but non-zero, for $T < T_g$. The entropy is then intensive in this temperature range, meaning that only a finite number of configurations, among the 2^N ones a priori available, are effectively occupied: the system is trapped in the lowest energy configurations.

References

1. R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New-York, 1975)
2. D.A. McQuarrie, *Statistical Mechanics* (Harper and Row, 1976)
3. T.L. Hill, *An Introduction to Statistical Thermodynamics* (Dover, 1986)
4. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford, 1987)
5. H. Goldstein, *Classical Mechanics*, 2nd edn (Addison-Wesley, 1980)
6. M. Le Bellac, *Quantum and Statistical Field Theory* (Oxford Science Publications, Oxford, 1992)
7. D.R. Nelson, M.E. Fisher, Soluble renormalization groups and scaling fields for low-dimensional Ising systems. *Ann. Phys.* **91**, 226 (1975)
8. S.F. Edwards, P.W. Anderson, Theory of spin glasses. *J. Phys. F* **5**, 965 (1975)
9. D. Sherrington, S. Kirkpatrick, Solvable model of a spin-glass. *Phys. Rev. Lett.* **35**, 1792 (1975)
10. M. Mézard, G. Parisi, M.A. Virasoro, *Spin-Glass Theory and Beyond* (World Scientific, Singapore, 1987)
11. D.C. Mattis, Solvable spin systems with random interactions. *Phys. Lett. A* **56**, 421 (1976)
12. B. Derrida, Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B* **24**, 2613 (1981)

Chapter 2

Non-stationary Dynamics and Stochastic Formalism

In the first part of this book, we have considered the stationary properties of physical systems composed of a large number of particles, using as a fundamental statistical object the joint distribution of all the degrees of freedom of the system (for instance positions and velocities, or spin variables). This steady state is expected to be reached after a transient regime, during which the macroscopic properties of the system evolve with time. Describing the statistical state of the system during this transient regime is also certainly of interest.

However, there is no known simple postulate (similar to the postulate of equiprobability of configurations having a given energy) to characterize the N -particle probability distribution in this time-dependent regime. Still, one can resort to the generic mathematical formalism of stochastic processes in order to describe statistically the time evolution of some specific variables of interest, like the position or velocity of a probe particle immersed in a fluid. This formalism is presented in Sect. 2.1, in the simplest case of Markov processes. The example of the random evolution of a single degree of freedom in a noisy environment (diffusive motion), leading to the Langevin and Fokker-Planck equations, is discussed respectively in Sects. 2.2 and 2.3. In addition, there exists situations in which this random evolution can be much faster or much slower than a priori expected, leading to anomalous diffusion. A brief account of scaling arguments allowing for a qualitative understanding of anomalous diffusion is given in Sect. 2.4. Generic issues regarding the convergence to equilibrium statistics in the framework of Markovian stochastic processes are presented in Sect. 2.5. Interestingly, the equilibrium distribution may not exist in some cases, leading to an endless relaxation called aging regime. An example of such a situation is also provided.

2.1 Markovian Stochastic Processes and Master Equation

2.1.1 Definition of Markovian Stochastic Processes

Let us start with some basic considerations on stochastic processes. For more advanced reading on this topic, we refer the reader for instance to Ref. [1]. Roughly speaking, a stochastic process is a dynamical process whose evolution is random, and depends on the presently occupied state and possibly on the history of the system.

Considering first a discrete time process ($t = 0, 1, 2, \dots$), with a finite number N of configurations C , we denote as $T(C_{t+1}|C_t, C_{t-1}, \dots, C_0)$ the probability for the process to jump to a new configuration C_{t+1} between times t and $t + 1$, given the whole history $(C_t, C_{t-1}, \dots, C_0)$. Note that C_{t+1} can a priori be any of the N possible configurations, including the configuration C_t itself. The transition probability $T(C_{t+1}|C_t, C_{t-1}, \dots, C_0)$ can be considered as a conditional probability, so that the following normalization condition holds

$$\sum_{C_{t+1}} T(C_{t+1}|C_t, C_{t-1}, \dots, C_0) = 1. \quad (2.1)$$

Such a stochastic process is said to be Markovian if the transition probability $T(C_{t+1}|C_t, C_{t-1}, \dots, C_0)$ depends only on the configuration C_t occupied at time t , and not on previously occupied configurations. In short, Markovian processes are said to be “memoryless”. The transition probability is then defined without explicit reference to time t . In the following, we denote as $T(C'|C)$ the transition probability from configuration C to configuration C' . This transition probability satisfies the normalization condition

$$\sum_{C'} T(C'|C) = 1. \quad (2.2)$$

The above definition of discrete time Markovian stochastic processes (also called Markov chains) can be rather straightforwardly extended to several other cases of practical importance. First, the number of discrete configurations can be infinite, and this case is recovered by taking the limit $N \rightarrow \infty$ in the above definition. If configurations are no longer discrete, but are defined by a continuous variable y , a probability density $\tilde{T}(y'|y)$ needs to be introduced, in such a way that $\tilde{T}(y'|y)dy'$ is the probability to choose a new configuration in the interval $[y', y' + dy']$, starting from a given configuration y . The equivalent of the normalization condition Eq. (2.2) now reads

$$\int_{-\infty}^{\infty} \tilde{T}(y'|y)dy' = 1. \quad (2.3)$$

Another generalization consists in replacing the discrete time steps by a continuous time evolution. Interestingly, continuous time dynamics can be obtained from the discrete time dynamics in the limit of a vanishing time step. Hence instead of using

a time step $\Delta t = 1$ as above, we now take an infinitesimal step $\Delta t = dt$. In order to obtain a meaningful limit when $dt \rightarrow 0$, the transition probabilities $T(C'|C)$ from configuration C to configuration C' have to scale with dt in the following way:

$$\begin{aligned} T(C'|C) &= W(C'|C) dt + \mathcal{O}(dt^2) \quad \text{if } C' \neq C, \\ T(C|C) &= 1 - \sum_{C' (\neq C)} W(C'|C) dt + \mathcal{O}(dt^2), \end{aligned} \quad (2.4)$$

where $W(C'|C)$ is independent of dt . In other words, the evolution of continuous time Markovian stochastic processes is characterized by transition rates $W(C'|C)$, such that $W(C'|C)dt$ is the probability for the process to go from configuration C to a new configuration C' in a time interval $[t, t + dt]$.

Finally, in the case of a continuous time process represented by a continuous variables y , a density of transition rate $w(y'|y)$ should be defined, in such a way that $w(y'|y)dy'dt$ is the probability for the process to reach a value in the interval $[y', y' + dy']$ at time $t + dt$, starting from a value y at time t .

Beyond formal definitions and calculations, it is also important to be able to simulate Markovian stochastic processes on a computer. A brief account of elementary simulation methods is provided in Appendix B.

2.1.2 Master Equation and Detailed Balance

The master equation describes the time evolution of the probability to occupy a given configuration. The simplest situation corresponds to discrete time and discrete configurations. The evolution of the probability $P_t(C)$ to occupy configuration C at time t is given by

$$P_{t+1}(C) = \sum_{C'} T(C|C') P_t(C'). \quad (2.5)$$

The probability $P_{t+1}(C)$ is thus simply a sum over all possible configurations C' of the probability to go from C' to C , weighted by the probability to occupy the configuration C' at time t . It is easy to check, by summing over all configurations C and using the normalization equation (2.2), that Eq. (2.5) conserves the normalization of the probability $P_t(C)$; namely, if $\sum_C P_t(C) = 1$, then $\sum_C P_{t+1}(C) = 1$.

For continuous configurations y , a density $p_t(y)$ has to be introduced (i.e., $p_t(y) dy$ is the probability that the configuration at time t belongs to the interval $[y, y + dy]$), and the evolution equation reads:

$$p_{t+1}(y) = \int_{-\infty}^{\infty} \tilde{T}(y|y') p_t(y') dy'. \quad (2.6)$$

The evolution of continuous time processes can be derived from this discrete time equation, using again the limit of a vanishing time step dt . Considering a continuous time process with discrete configurations, we denote as $P(C, t)$ the probability to be in configuration C at time t . Combining Eqs. (2.4) and (2.5), we get

$$P(C, t + dt) = \sum_{C'(\neq C)} W(C|C')dt P(C', t) + \left(1 - \sum_{C'(\neq C)} W(C'|C)dt\right) P(C, t). \quad (2.7)$$

Expanding the left-hand-side of this last equation to first order in dt , as

$$P(C, t + dt) = P(C, t) + \frac{dP}{dt}(C, t) dt + \mathcal{O}(dt^2) \quad (2.8)$$

we eventually find, in the limit $dt \rightarrow 0$, that the probability $P(C, t)$ evolves according to the master equation:

$$\frac{dP}{dt}(C, t) = -P(C, t) \sum_{C'(\neq C)} W(C'|C) + \sum_{C'(\neq C)} W(C|C')P(C', t). \quad (2.9)$$

The first term in the right-hand-side can be interpreted as a “loss” term (i.e., the sum of all the possibilities to exit configuration C), while the second term can be thought of as a “gain” term (the sum of all the possibilities to arrive at configuration C , starting from any other configuration). A similar equation is obtained in the case of continuous configurations y for the probability density $p(y, t)$, by basically replacing discrete sums by integrals in Eq. (2.9):

$$\frac{\partial p}{\partial t}(y, t) = -p(y, t) \int_{-\infty}^{\infty} dy' w(y'|y) + \int_{-\infty}^{\infty} dy' w(y|y') p(y', t). \quad (2.10)$$

Further generalization to multidimensional configurations $\mathbf{y} = (y_1, \dots, y_n)$ is also straightforward. From now on, we will work mainly with discrete configurations as far as formal and generic calculations are concerned, keeping in mind that the continuous variable case can be obtained by switching from discrete to continuous notations.

An interesting property of continuous time master equations is the notion of detailed balance, which is related to the steady-state (i.e., time-independent) solution of the master equation. From Eq. (2.9), a time-independent solution $P(C)$ satisfies, for all configurations C

$$\sum_{C'(\neq C)} [-W(C'|C)P(C) + W(C|C')P(C')] = 0. \quad (2.11)$$

It may happen, for some specific stochastic processes, that the term between bracket vanishes for all C' , namely

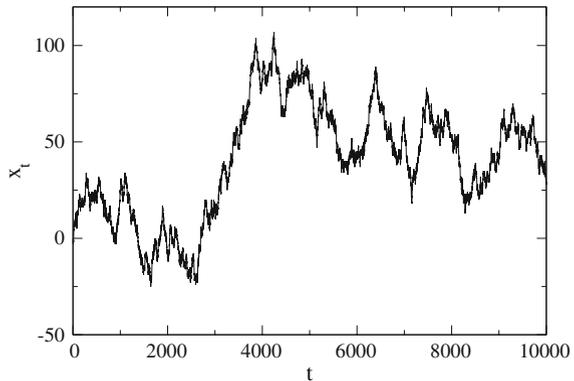
$$\forall(C, C'), \quad W(C'|C)P(C) = W(C|C')P(C'). \quad (2.12)$$

This situation is referred to as detailed balance. Processes satisfying detailed balance are much easier to handle analytically. Besides this practical advantage, detailed balance also plays an important role in the stochastic modeling of microscopic physical processes (i.e., at the molecular scale). This is due to the fact that detailed balance can be interpreted as the stochastic counterpart of the microreversibility property satisfied by the Hamiltonian dynamics—see Sect. 1.1. Indeed, the probability to observe, once a statistical steady-state is reached, an elementary trajectory from C at time t to C' at time $t + dt$ is $W(C'|C) dt P(C)$, while the probability to observe the reverse trajectory is $W(C|C') dt P(C')$. The equality of these two probabilities, to be thought of as a statistical microreversibility, precisely yields the detailed balance relation (2.12). Hence in order to model, at a coarse-grained level, the dynamics of a microscopic physical system through a Markovian stochastic process, it is natural to assume that the process satisfies detailed balance (in addition to the appropriate conservation laws, like energy conservation).

2.1.3 A Simple Example: The One-Dimensional Random Walk

A simple and illustrative example of stochastic process is the one-dimensional random walk, where a “particle” moves at random on a one-dimensional lattice. Let us consider first the discrete time case: a particle can take only discrete positions $x = \dots, -2, -1, 0, 1, 2, \dots$ on a line. Between times t and $t + 1$, the particle randomly jumps to one of the two neighboring sites, so that $x_{t+1} = x_t + \epsilon_t$, with $\epsilon_t = \pm 1$ with equal probabilities. The random variables ϵ_t and $\epsilon_{t'}$, with $t \neq t'$, are independent and identically distributed. An illustration of a random walk is given in Fig. 2.1.

Fig. 2.1 Illustration of a random walk



The average value and the variance of this process can be derived straightforwardly. We first note that $\langle x_{t+1} \rangle = \langle x_t \rangle$, so that $\langle x_t \rangle = \langle x_0 \rangle$ for all t (the notation $\langle \dots \rangle$ denotes an ensemble average, that is an average over a very large number of samples of the same process; it may thus depend on time). For instance, if the walk starts with probability 1 from $x_0 = 0$, then all subsequent averages $\langle x_t \rangle = 0$.

Let us now compute the variance of the process, defined as

$$\text{Var}(x_t) = \langle x_t^2 \rangle - \langle x_t \rangle^2. \quad (2.13)$$

We assume for simplicity that $\langle x_t \rangle = 0$, so that $\text{Var}(x_t) = \langle x_t^2 \rangle$ (the generalization to $\langle x_t \rangle \neq 0$ is however straightforward). From $x_{t+1} = x_t + \epsilon_t$, we get

$$x_{t+1}^2 = x_t^2 + 2x_t\epsilon_t + 1, \quad (2.14)$$

taking into account that $\epsilon_t^2 = 1$. Computing the ensemble average of Eq. (2.14) yields

$$\langle x_{t+1}^2 \rangle = \langle x_t^2 \rangle + 2\langle x_t \rangle \langle \epsilon_t \rangle + 1, \quad (2.15)$$

using the fact that x_t depends only on $\epsilon_{t'}$ with $t' < t$, so that x_t and ϵ_t are independent random variables. As $\langle \epsilon_t \rangle = 0$, it follows that $\langle x_{t+1}^2 \rangle = \langle x_t^2 \rangle + 1$, so that $\langle x_t^2 \rangle = \langle x_0^2 \rangle + t$. If $x_0 = 0$ with probability 1, one has $\langle x_0^2 \rangle = 0$, and $\langle x_t^2 \rangle = t$. This means that the typical position reached by the walk after t steps is of the order of \sqrt{t} .

The present random walk problem bears a direct relationship to the Central Limit Theorem [2, 3]—see Chap. 6. As the position x_t of the random walk can be expressed as $x_t = \sum_{t'=0}^{t-1} \epsilon_{t'}$, where $(\epsilon_0, \dots, \epsilon_{t-1})$ are independent and identically distributed random variables, the distribution of the position of the random walk can be approximated for a large time t , using the Central Limit Theorem, as

$$P(x, t) \approx \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}. \quad (2.16)$$

Alternatively, one may endow the random walk problem with a continuous time dynamics. Labeling with an integer n the sites of the lattice, the transition rate $W(n'|n)$ from site n to site n' is given by

$$W(n'|n) = \begin{cases} \frac{\nu}{2} & \text{if } n' = n \pm 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.17)$$

where ν is a characteristic frequency (the inverse of a time scale) of the process. The master equation reads

$$\frac{dP_n}{dt} = - \sum_{n'(\neq n)} W(n'|n) P_n(t) + \sum_{n'(\neq n)} W(n|n') P_{n'}(t). \quad (2.18)$$

Replacing the transition rates by their expression given in Eq. (2.17), one finds

$$\frac{dP_n}{dt} = -\nu P_n(t) + \frac{\nu}{2} P_{n+1}(t) + \frac{\nu}{2} P_{n-1}(t). \quad (2.19)$$

The evolution of the probability distribution $P_n(t)$ can be evaluated from Eq. (2.19), for instance by integrating it numerically. However, one may be interested in making analytical predictions in the large time limit, and such a discrete-space equation is not easy to handle in this case. To this aim, it is thus useful to use a procedure called “continuous limit”, through which the discrete-space equation (2.19) can be approximated by a partial differential equation. To be more specific, let us call a the lattice spacing (which was set above to $a = 1$). At large time $t \gg 1/\nu$, the distribution $P_n(t)$ is expected to vary over spatial scales much larger than the lattice spacing a ; in other words, one has

$$|P_{n+1}(t) - P_n(t)| \ll P_n(t). \quad (2.20)$$

Plotting $P_n(t)$ as a function of space, it thus appears essentially continuous. Hence we postulate the existence of a distribution $p(x, t)$ of the continuous variable x , such that the discrete-space distribution can be approximated as $P_n(t) \approx a p(na, t)$. The prefactor a is included to ensure a correct normalization, $\sum_n P_n(t) = 1$. Indeed, one has for $a \rightarrow 0$

$$\sum_n P_n(t) = a \sum_n p(na, t) \rightarrow \int_{-\infty}^{\infty} p(x, t) dx. \quad (2.21)$$

For consistency, it is thus necessary to assume that $p(x, t)$ is normalized such that $\int_{-\infty}^{\infty} p(x, t) dx = 1$.

Replacing $P_n(t)$ by $a p(na, t)$ in the master equation (2.19), one obtains

$$\frac{\partial p}{\partial t}(x, t) = -\nu p(x, t) + \frac{\nu}{2} p(x + a, t) + \frac{\nu}{2} p(x - a, t). \quad (2.22)$$

As a is small, one can expand $p(x \pm a, t)$ to second order in a , leading to

$$p(x \pm a, t) = p(x, t) \pm a \frac{\partial p}{\partial x}(x, t) + \frac{a^2}{2} \frac{\partial^2 p}{\partial x^2}(x, t) + \mathcal{O}(a^3). \quad (2.23)$$

The linear terms in a appearing in Eq. (2.22) cancel out, so that this equation reduces to

$$\frac{\partial p}{\partial t}(x, t) = \frac{\nu a^2}{2} \frac{\partial^2 p}{\partial x^2}(x, t) \quad (2.24)$$

which is called the diffusion equation. This equation appears in numerous problems in physics, like the diffusion of heat in a material, or the diffusion of dye in water for instance. The coefficient $\frac{1}{2}\nu a^2$ has to take a finite value $D > 0$ for the equation to be well-defined. As the lattice spacing a goes to zero, it is thus necessary that ν

simultaneously goes to infinity, which means that the 'microscopic' process appears very fast on the scale of observation.

Equation (2.24) has several simple solutions of interest. For instance, if the diffusing particle is bound to stay on a segment $[-L, L]$, the long-time limit distribution is a flat and time-independent distribution over the segment, $p(x) = (2L)^{-1}$. In other words, diffusion tends to flatten, or smoothen, the distribution. In contrast, if the particle can diffuse on the entire line without bound, the distribution $p(x, t)$ never reaches a steady-state regime, but rather enters a scaling regime in which the distribution keeps broadening with time, with a well-defined Gaussian shape:

$$p(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt}. \quad (2.25)$$

Note the analogy with the result obtained from the Central Limit Theorem in the discrete time case—see Eq. (2.16).

2.2 Langevin Equation

2.2.1 Phenomenological Approach

The above random walk example was quite simple to investigate, but had little explicit connection with physical systems. We now present another standard example based on a physical phenomenology. Let us imagine a probe particle immersed in a fluid, such that the size of the particle is small at the macroscopic scale, but still much larger than the typical size of the molecules of the fluid. For the sake of simplicity, we restrict the presentation to a one-dimensional system, but the more realistic three-dimensional situation would follow the same line.

We choose the mass of the probe particle as the unit mass. The acceleration of the particle is then governed by the force F_{coll} exerted by the collisions with the other particles:

$$\frac{dv}{dt} = F_{\text{coll}}, \quad (2.26)$$

where v is the velocity of the probe particle. Since the collisional force F_{coll} is strongly fluctuating, the basic idea is to decompose it into a (velocity-dependent) average force, and a purely fluctuating (or noise) part:

$$F_{\text{coll}} = \langle F_{\text{coll}} \rangle|_v + \xi(t). \quad (2.27)$$

Here, the average force $\langle F_{\text{coll}} \rangle|_v$ is computed as an average over a large number of samples of the process, conditioned to a given value v of the velocity. By definition,

the noise $\xi(t)$ has zero mean, $\langle \xi(t) \rangle = 0$. In principle, the statistics of the noise could also depend on the velocity v . We assume here that it is independent of v , but we will come back to this point below. To proceed further, it is necessary to choose a specific model for both $\langle F_{\text{coll}} \rangle$ and $\xi(t)$. The average force $\langle F_{\text{coll}} \rangle$ can be interpreted as an effective friction force, which slows down the probe particle; it is thus natural to choose, as a first approximation, a linear friction force $\langle F_{\text{coll}} \rangle = -\gamma v$, with $\gamma > 0$ a friction coefficient.

Then, a model of the noise should be given. Beside the property $\langle \xi(t) \rangle = 0$, its two-time correlation should be specified. Intuitively, one expects collisions occurring at different times to be essentially uncorrelated, so that one should have $\langle \xi(t) \xi(t') \rangle = 0$ for $|t - t'| \gg \tau_{\text{col}}$, where τ_{col} is the typical duration of a collision. Taking into account time-translation invariance, the correlation function of $\xi(t)$ may thus be written as

$$\langle \xi(t) \xi(t') \rangle = C(t - t'), \quad (2.28)$$

where $C(u)$ is an even function, that decays on a characteristic time scale τ_{col} , and converges rapidly to zero when $|u| \rightarrow \infty$. In practice, the dynamics of the velocity v of the probe particle occurs on a time scale γ^{-1} that is much larger than τ_{col} . In this limit, the only quantity that plays a role in the dynamics of v is the integral of the correlation function, that we denote as Γ ,

$$\Gamma = \int_{-\infty}^{\infty} C(t) dt. \quad (2.29)$$

Since the detailed shape of $C(t)$ plays no role in this limit, it is convenient to replace $C(t)$ by a delta function $\Gamma \delta(t - t')$, keeping the same value of the integral (a basic introduction to Dirac delta function can be found in Appendix A).

Altogether, Eq. (2.26) can be rewritten as:

$$\frac{dv}{dt} = -\gamma v + \xi(t), \quad (2.30)$$

with

$$\langle \xi(t) \rangle = 0, \quad \langle \xi(t) \xi(t') \rangle = \Gamma \delta(t - t'). \quad (2.31)$$

Such an equation is called a linear Langevin equation, with additive white noise. The equation is called linear because the deterministic part of the dynamics, the term $-\gamma v$, is linear with respect to the variable v . Additive noise simply means that the noise is introduced in the equation as an additive term which is independent of the variable v . We will see in Sect. 2.2.3 more complicated situations where the noise term is coupled to the dynamical variable.

2.2.2 Basic Properties of the Linear Langevin Equation

We now study some elementary properties of the linear Langevin equation (2.30), namely the ensemble averages $\langle v(t) \rangle$ and $\langle v(t)^2 \rangle$. For simplicity, we take as initial condition a fixed value $v(0) = v_0$. We first note, computing the ensemble average of Eq. (2.30):

$$\frac{d}{dt} \langle v(t) \rangle = -\gamma \langle v(t) \rangle, \quad (2.32)$$

that the ensemble-averaged velocity $\langle v(t) \rangle$ obeys the same equation as the non-averaged velocity, except that noise is now absent. This property is specific to the linear Langevin equation, and would not be present if we had included a non-linear dependence on v in the friction force—e.g., $\langle F_{\text{coll}} \rangle = -\gamma v - \gamma_3 v^3$. The solution of Eq. (2.32) is a decaying exponential:

$$\langle v(t) \rangle = v_0 e^{-\gamma t}. \quad (2.33)$$

More interestingly, the effect of the noise has a deep impact on the evolution of the variance of the velocity, $\text{Var}[v(t)] = \langle v(t)^2 \rangle - \langle v(t) \rangle^2$. In order to compute $\langle v(t)^2 \rangle$, we first determine the explicit time-dependence of $v(t)$, considering $\xi(t)$ as an arbitrary given function. Following standard mathematical methods, the general solution of Eq. (2.30) is given by the sum of the general solution of the homogeneous equation (i.e., the noiseless equation) and of a particular solution of the full equation. The general solution of the homogeneous equation is $v_h(t) = A_0 e^{-\gamma t}$, where A_0 is an arbitrary constant. In order to determine a particular solution, one can use the so-called “variation of the constant” method, which indicates that such a solution should be looked for in the form $v_p(t) = A(t) e^{-\gamma t}$, that is, simply replacing the constant A_0 in the solution $v_h(t)$ of the homogeneous equation by a function $A(t)$ to be determined. Inserting $v_p(t)$ in Eq. (2.30), we get

$$\frac{dA}{dt} e^{-\gamma t} = \xi(t) \quad (2.34)$$

whence the solution

$$A(t) = \int_0^t e^{\gamma t'} \xi(t') dt' \quad (2.35)$$

follows—since we look for a particular solution at this stage, there is no need to add a constant term to Eq. (2.35). Altogether, one finds for $v(t)$, taking into account the initial condition $v(0) = v_0$,

$$v(t) = v_0 e^{-\gamma t} + e^{-\gamma t} \int_0^t e^{\gamma t'} \xi(t') dt'. \quad (2.36)$$

Computing $v(t)^2$ yields

$$v(t)^2 = v_0^2 e^{-2\gamma t} + e^{-2\gamma t} \left(\int_0^t e^{\gamma t'} \xi(t') dt' \right)^2 + 2v_0 e^{-2\gamma t} \int_0^t e^{\gamma t'} \xi(t') dt'. \quad (2.37)$$

Now taking an ensemble average, the last term vanishes because $\langle \xi(t) \rangle = 0$, and we get

$$\langle v(t)^2 \rangle = v_0^2 e^{-2\gamma t} + \left\langle e^{-2\gamma t} \left(\int_0^t e^{\gamma t'} \xi(t') dt' \right)^2 \right\rangle. \quad (2.38)$$

The first term on the right hand side of Eq. (2.38) is precisely $\langle v(t) \rangle^2$, so that

$$\text{Var}[v(t)] = \left\langle e^{-2\gamma t} \left(\int_0^t e^{\gamma t'} \xi(t') dt' \right)^2 \right\rangle. \quad (2.39)$$

The square of the integral can be expanded as a product of two integrals, which in turn can be converted into a double integral:

$$\begin{aligned} \left(\int_0^t e^{\gamma t'} \xi(t') dt' \right)^2 &= \int_0^t e^{\gamma t'} \xi(t') dt' \int_0^t e^{\gamma t''} \xi(t'') dt'' \\ &= \int_0^t dt' \int_0^t dt'' e^{\gamma(t'+t'')} \xi(t') \xi(t'') \end{aligned} \quad (2.40)$$

so that Eq. (2.39) eventually turns into

$$\text{Var}[v(t)] = e^{-2\gamma t} \int_0^t dt' \int_0^t dt'' e^{\gamma(t'+t'')} \langle \xi(t') \xi(t'') \rangle \quad (2.41)$$

(we recall that the ensemble average can be interverted with linear operations like integrals or derivatives). Using the expression (2.31) of $\langle \xi(t') \xi(t'') \rangle$, we get

$$\text{Var}[v(t)] = \Gamma e^{-2\gamma t} \int_0^t dt' \int_0^t dt'' e^{\gamma(t'+t'')} \delta(t' - t''). \quad (2.42)$$

It is useful to make a change of variable here, replacing t'' by the variable $y = t'' - t'$ in the second integral, which yields

$$\text{Var}[v(t)] = \Gamma e^{-2\gamma t} \int_0^t dt' e^{2\gamma t'} \int_{-t'}^{t-t'} dy e^{\gamma y} \delta(y). \quad (2.43)$$

The second integral in Eq. (2.43) can easily be computed, thanks to the properties of the delta function, leading to

$$\int_{-t'}^{t-t'} dy e^{\gamma y} \delta(y) = 1. \quad (2.44)$$

We thus eventually find, after integration of Eq. (2.43),

$$\text{Var}[v(t)] = \frac{\Gamma}{2\gamma} (1 - e^{-2\gamma t}). \quad (2.45)$$

Hence the variance starts from a zero value at $t = 0$ (the value v_0 at $t = 0$ is non-random), and progressively grows until reaching the asymptotic limit $\Gamma/(2\gamma)$. As $\langle v(t) \rangle \rightarrow 0$ when $t \rightarrow \infty$, the variance reduces to $\langle v^2 \rangle$ at large time, and this value can be identified with the equilibrium average. It is known from equilibrium statistical physics (see Sect. 1.2.4) that $\langle \frac{1}{2} v^2 \rangle_{\text{eq}} = \frac{1}{2} k_B T$ (equipartition relation), where T is the temperature of the surrounding liquid, and k_B the Boltzmann constant—we recall that the mass of the probe particle was set to unity.¹ Hence equilibrium statistical physics imposes a relation between the two phenomenologically introduced coefficients Γ and γ , namely $\Gamma = 2\gamma k_B T$.

In addition, using a slight generalization of the above calculation, it is also straightforward to show that for large times ($t, t' \gg \gamma^{-1}$), the correlation of v decays exponentially,

$$\langle v(t)v(t') \rangle = \frac{\Gamma}{2\gamma} e^{-\gamma|t-t'|}, \quad (2.46)$$

which reduces to the infinite time limit of Eq. (2.45) for $t = t'$.

Before considering more general forms of Langevin equations, it is useful to say a word on the practical way to integrate numerically the Langevin equation (2.30). Similarly to ordinary differential equations, one needs to discretize the equation using small time steps Δt . This discretization is obtained by computing the integral of Eq. (2.30) over a time interval $[t, t + \Delta t]$, yielding

$$v(t + \Delta t) = v(t) - \gamma v(t) \Delta t + \int_t^{t+\Delta t} \xi(t') dt' \quad (2.47)$$

with the approximation $\int_t^{t+\Delta t} v(t') dt' \approx v(t) \Delta t$, valid for small Δt . From Eq. (2.31), the quantity $\Delta W \equiv \int_t^{t+\Delta t} \xi(t') dt'$ is a random variable with zero mean and variance

$$\langle (\Delta W)^2 \rangle = \int_t^{t+\Delta t} dt' \int_t^{t+\Delta t} dt'' \langle \xi(t') \xi(t'') \rangle = \Gamma \Delta t. \quad (2.48)$$

¹Reintroducing the mass m , the equipartition relation reads $\langle \frac{1}{2} m v^2 \rangle_{\text{eq}} = \frac{1}{2} k_B T$.

From its definition, ΔW can be interpreted as a sum of a very large number of statistically independent contributions, so that it is natural to assume that the distribution of ΔW is Gaussian from the Central Limit Theorem—see Chap. 6. Hence we assume that ΔW is a Gaussian random variable with zero mean and variance $\Gamma \Delta t$ (more rigorous justifications can be found, e.g., in Ref. [4]). Discretized stochastic trajectories can thus be numerically obtained from Eq. (2.47), drawing at each time step a new random value of ΔW (values of ΔW at different time steps are statistically independent). Time-dependent average values of observables are then obtained by averaging a given observable (e.g., $v(t)$ or $v(t)^2$) over many independent stochastic trajectories with the same initial condition.

2.2.3 More General Forms of the Langevin Equation

We have studied in the previous section the simplest version of the Langevin equation, namely the linear Langevin equation with additive noise. We now wish to briefly mention several important generalizations of this equation. From now on, we will generically call x the variable evolving according to the Langevin equation. It may be any type of ‘mesoscopic’ physical observable, which is sensitive to the presence of fluctuations, but evolves on times scales that are sufficiently large for a Langevin type of description to be relevant. In the example of the probe particle discussed in Sect. 2.2.1, this means that the probe particle is much heavier than the molecules it collides with, but still much lighter than, say, a grain of sand which would not feel any fluctuations in the force exerted by the surrounding fluid (if the fluid is at rest).

A first generalization of the linear Langevin equation is to consider a non-linear deterministic term in the equation:

$$\frac{dx}{dt} = Q(x) + \xi(t) \quad (2.49)$$

where $Q(x)$ is an arbitrary function of x . The white noise $\xi(t)$ still satisfies $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \Gamma \delta(t-t')$. In the following subsections, we will generically consider this case when dealing with the Langevin equation. More generally, one may consider Langevin equations coupling an arbitrary number N of variables x_i ,

$$\frac{dx_i}{dt} = Q_i(x_1, \dots, x_N) + \xi_i(t), \quad i = 1, \dots, N, \quad (2.50)$$

where the N stochastic variables $\xi_i(t)$ are white noises with correlations

$$\langle \xi_i(t)\xi_j(t') \rangle = \Gamma_{ij} \delta(t-t'), \quad i, j = 1, \dots, N. \quad (2.51)$$

Another important generalization of the Langevin equation is the one with multiplicative noise, in contrast to additive noise. For a single variable, the Langevin

equation with multiplicative noise takes the generic form

$$\frac{dx}{dt} = Q(x) + B(x) \xi(t) \quad (2.52)$$

with a white noise $\xi(t)$ of unit amplitude, $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$ (the amplitude Γ previously considered can be reabsorbed into the function B). In this case, the amplitude of the noise term $B(x) \xi(t)$ depends on the value of the variable x . This is important for instance in the modeling of absorbing phase transitions (see Chap. 4) where fluctuations vanish in the absorbing state, when there are no more particles in the system. The specificity of the multiplicative Langevin equation (2.52) is that it requires the specification of a discretization scheme in order to be well-defined, as different discretization schemes lead to different results. A rigorous description of Eq. (2.52) can be achieved within the mathematical framework of stochastic calculus [4]. Here, we however stick to a heuristic viewpoint and simply describe the two main interpretations of Eq. (2.52) at an elementary level. From a mathematical perspective, the most natural interpretation is the Ito one, corresponding to the following discretization of Eq. (2.52). Introducing an increasing sequence of discrete times $t_i = i\Delta t$, the Ito interpretation corresponds to

$$x(t_{i+1}) = x(t_i) + Q(x(t_i)) \Delta t + B(x(t_i)) \Delta W_i \quad (2.53)$$

where ΔW_i is a Gaussian random variable with zero mean and variance Δt . An alternative interpretation, commonly used in the physics community, is the Stratonovich one, which is expressed in terms of the discretized equation

$$x(t_{i+1}) = x(t_i) + Q(x(t_i)) \Delta t + B\left(\frac{x(t_i) + x(t_{i+1})}{2}\right) \Delta W_i. \quad (2.54)$$

Both formulations lead to consistent interpretations of Eq. (2.52). The choice of one or the other may in some cases be related to the problem at hand. For instance, starting from a problem with a finite, but small correlation time of the noise, the correct interpretation in the white noise limit is the Stratonovich one [4].

Finally, another type of generalization consists in changing the properties of the noise, assuming that the noise has a finite correlation time. A typical case is that of an exponentially decaying correlation function

$$\langle \xi(t)\xi(t') \rangle = \frac{\Gamma}{2\tau} e^{-|t-t'|/\tau} \quad (2.55)$$

which reduces to a white noise in the limit $\tau \rightarrow 0$. For a non-zero τ , such a noise is often called a ‘colored noise’, and it may be obtained for instance from an underlying Langevin equation with white noise, as illustrated in Eq. (2.46). Then an equation like

$$\frac{dv}{dt} = -\gamma v + \xi(t), \quad (2.56)$$

yields for the mean square velocity in the limit $t \rightarrow \infty$, generalizing the calculation made in Sect. 2.2.1,

$$\langle v(t)^2 \rangle = \frac{\Gamma}{2\gamma(1 + \gamma\tau)}. \quad (2.57)$$

As discussed above, white noise is often used to model equilibrium systems, and the ratio Γ/γ is in this case related to temperature. In contrast, colored noise is relevant when modeling out-of-equilibrium systems, and no generic relation between Γ , γ , τ and macroscopic parameters of the problem (like temperature in the equilibrium case) is known.

2.2.4 Relation to Random Walks

After having introduced the Langevin equation from a physical perspective (that of a probe particle immersed in a fluid), it is interesting to present the Langevin equation from another perspective, that of random walks. To this aim, we come back to the random walk model introduced in Sect. 2.1.3 and generalize it by including a small bias in the displacements. We consider a discrete time dynamics with a time step Δt , and we call a the lattice spacing. At time $t + \Delta t$, the new position $x_{t+\Delta t}$ is chosen according to $x_{t+\Delta t} = x_t + \epsilon_t$, where ϵ_t is given by

$$\epsilon_t = \begin{cases} a & \text{with prob. } \frac{\nu}{2}(1 + aq(x_t))\Delta t, \\ -a & \text{with prob. } \frac{\nu}{2}(1 - aq(x_t))\Delta t, \\ 0 & \text{with prob. } 1 - \nu\Delta t. \end{cases} \quad (2.58)$$

Note that the above dynamical rules can be interpreted as a discretized version of a continuous time dynamics, as seen from the presence of the time step Δt and from the allowed value $\epsilon_t = 0$. Let us define $\Delta x_t \equiv x_{t+\Delta t} - x_t$. The dynamical rules $x_{t+\Delta t} = x_t + \epsilon_t$ can be rewritten as

$$\frac{\Delta x_t}{\Delta t} = \frac{\epsilon_t}{\Delta t} \quad (2.59)$$

which is the analog of Eq. (2.26), provided that x_t is interpreted as a velocity; $\epsilon_t/\Delta t$ then plays the role of a random force. Computing the average value of this ‘force’, we find using Eq. (2.58)

$$\left\langle \frac{\epsilon_t}{\Delta t} \right\rangle = a^2 \nu q(x_t). \quad (2.60)$$

Note that the average is taken over ϵ_t , for a fixed value of x_t . Let us now consider the fluctuating part of the ‘force’, and define

$$\xi_t = \frac{1}{\Delta t} (\epsilon_t - \langle \epsilon_t \rangle), \quad (2.61)$$

which is thus the discrete-time analog of $\xi(t)$ introduced in Sect. 2.2.1. We wish to evaluate the correlation of ξ_t , given by

$$\langle \xi_t \xi_{t'} \rangle = \frac{1}{(\Delta t)^2} \left\langle (\epsilon_t - \langle \epsilon_t \rangle) (\epsilon_{t'} - \langle \epsilon_{t'} \rangle) \right\rangle. \quad (2.62)$$

For $t \neq t'$, $\langle \xi_t \xi_{t'} \rangle$ is thus equal to zero, as ϵ_t and $\epsilon_{t'}$ are independent random variables. If $t = t'$, one has $\langle \xi_t \xi_{t'} \rangle = \text{Var}(\epsilon_t) / (\Delta t)^2$. Introducing (k, k') through $t = k\Delta t$ and $t' = k'\Delta t$, Eq. (2.62) reads

$$\langle \xi_t \xi_{t'} \rangle = \frac{1}{(\Delta t)^2} \text{Var}(\epsilon_t) \delta_{k,k'} \quad (2.63)$$

where $\delta_{k,k'}$ is the Kronecker delta symbol, equal to 1 if $k = k'$ and to zero otherwise. Evaluating the variance of ϵ_t , we find

$$\text{Var}(\epsilon_t) = a^2 \nu \Delta t + \mathcal{O}(\Delta t^2), \quad (2.64)$$

so that to leading order in Δt ,

$$\langle \xi_t \xi_{t'} \rangle = a^2 \nu \frac{\delta_{k,k'}}{\Delta t}. \quad (2.65)$$

This expression is the analog of Eq. (2.31), and the role played by τ_{col} in the physical approach to the Langevin equation (see Sect. 2.2.1) is now played by Δt . To provide further evidence for this correspondence, we point out that $\delta_{k,k'} / \Delta t$ can be interpreted as the discretized version of the Dirac distribution. Indeed, from the definition of the Kronecker delta symbol, one can write for an arbitrary function f

$$\sum_{k'=-\infty}^{\infty} \Delta t f(k' \Delta t) \frac{\delta_{k,k'}}{\Delta t} = f(k \Delta t), \quad (2.66)$$

which is precisely the discretized version of the fundamental property (A.1) of the Dirac delta function. Hence taking the limit $\Delta t \rightarrow 0$ and $a \rightarrow 0$, one can reformulate the above biased random walk problem as a Langevin equation, namely

$$\frac{dx}{dt} = Q(x) + \xi(t) \quad (2.67)$$

where $Q(x) \equiv a^2 \nu q(x)$, and where the noise $\xi(t)$ satisfies

$$\langle \xi(t) \rangle = 0, \quad \langle \xi(t) \xi(t') \rangle = \Gamma \delta(t - t') \text{ with } \Gamma = a^2 \nu. \quad (2.68)$$

2.3 Fokker-Planck Equation

The Fokker-Planck equation describes the evolution of the probability distribution $p(x, t)$ of a variable x obeying a Langevin equation. It can be derived in several ways, one of the simplest being to start from the above biased random walk problem, and to derive the continuous limit of the master equation, following the same lines as for the derivation of the diffusion equation—see Sect. 2.1.3.

2.3.1 Continuous Limit of a Discrete Master Equation

Starting from the biased random walk model of Sect. 2.2.4, we consider the continuous time version of the model, and write the corresponding transition rates $W(n'|n)$, where $n = x/a$ is an integer labeling the sites of the one-dimensional lattice:

$$W(n'|n) = \begin{cases} \frac{\nu}{2}(1 + aq_n) & \text{if } n' = n + 1 \\ \frac{\nu}{2}(1 - aq_n) & \text{if } n' = n - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (2.69)$$

To lighten notations, we have denoted $q(na)$ as q_n . Formally, one can write the transition rates as

$$W(n'|n) = \frac{\nu}{2}(1 + aq_n)\delta_{n',n+1} + \frac{\nu}{2}(1 - aq_n)\delta_{n',n-1}. \quad (2.70)$$

The master equation then reads

$$\frac{dP_n}{dt} = -\nu P_n(t) + \frac{\nu}{2}(1 + aq_{n-1})P_{n-1}(t) + \frac{\nu}{2}(1 - aq_{n+1})P_{n+1}(t). \quad (2.71)$$

We now take the continuous limit of this master equation. Writing, as in Sect. 2.1.3, $P_n(t) = ap(na, t)$, where $p(x, t)$ is a distribution of the continuous variable x satisfying $\int_{-\infty}^{\infty} p(x, t) dx = 1$, we have

$$\frac{\partial p}{\partial t}(x, t) = -\nu p(x, t) + \frac{\nu}{2}[1 + aq(x-a)]p(x-a, t) + \frac{\nu}{2}[1 - aq(x+a)]p(x+a, t). \quad (2.72)$$

Expanding $p(x \pm a, t)$ and $q(x \pm a)$ to second order in a , we get

$$p(x \pm a, t) = p(x, t) \pm a \frac{\partial p}{\partial x}(x, t) + \frac{a^2}{2} \frac{\partial^2 p}{\partial x^2}(x, t) + \mathcal{O}(a^2), \quad (2.73)$$

$$q(x \pm a) = q(x) \pm a q'(x) + \frac{a^2}{2} q''(x) + \mathcal{O}(a^2). \quad (2.74)$$

Gathering results, one then finds, keeping only terms up to order a^2 in Eq. (2.72):

$$\frac{\partial p}{\partial t}(x, t) = -a^2 \nu q(x) \frac{\partial p}{\partial x} - a^2 \nu q'(x) p(x, t) + \frac{a^2 \nu}{2} \frac{\partial^2 p}{\partial x^2}. \quad (2.75)$$

We note that $a^2 \nu$ is related both to the diffusion coefficient D introduced in Sect. 2.1.3, and to the coefficient Γ characterizing the correlation of the noise in Sect. 2.2.4:

$$a^2 \nu = 2D = \Gamma. \quad (2.76)$$

In order to have a well-defined continuous limit, one must here again take the limits $a \rightarrow 0$ and $\nu \rightarrow \infty$ in such a way that $a^2 \nu$ converges to a finite value. Defining $Q(x) = \Gamma q(x)$, Eq. (2.75) can be rewritten as

$$\frac{\partial p}{\partial t}(x, t) = -\frac{\partial}{\partial x} \left(Q(x) p(x, t) \right) + \frac{\Gamma}{2} \frac{\partial^2 p}{\partial x^2}. \quad (2.77)$$

This equation is called a Fokker-Planck equation. It describes, from another perspective, the same random process as the Langevin equation (2.67).

As an example of application of the Fokker-Planck equation, we come back to the probe particle studied in Sect. 2.2.1. In this case, the variable x is replaced by the velocity v , and the bias function is given by $Q(v) = -\gamma v$. The Fokker-Planck equation reads

$$\frac{\partial p}{\partial t}(v, t) = \gamma \frac{\partial}{\partial v} \left(v p(v, t) \right) + \frac{\Gamma}{2} \frac{\partial^2 p}{\partial v^2}, \quad (2.78)$$

where the coefficients Γ and γ are related through $\Gamma = 2\gamma k_B T$. It can be checked that the solution of this equation, with initial condition $p(v, t = 0) = \delta(v - v_0)$ —i.e., the initial velocity is non-random and equal to v_0 —is given by

$$p(v, t) = \left[2\pi k_B T (1 - e^{-2\gamma t}) \right]^{-1/2} \exp \left[-\frac{(v - v_0 e^{-\gamma t})^2}{2k_B T (1 - e^{-2\gamma t})} \right]. \quad (2.79)$$

One can check that the mean velocity $\langle v \rangle$ and the variance $\text{Var}[v(t)]$ correspond to the ones calculated from the Langevin equation—see Eqs. (2.33) and (2.45). This process, namely a random walk confined by a quadratic potential, is also called Ornstein-Uhlenbeck process.

2.3.2 Kramers-Moyal Expansion

More generally, the Fokker-Planck equation may be derived from an arbitrary master equation provided the random variable performs only small jumps. Let us consider a

stochastic Markov process defined by transition rates $W(x'|x)$. The master equation reads

$$\frac{\partial p}{\partial t}(x, t) = \int dx' [W(x|x')p(x', t) - W(x'|x)p(x, t)] \quad (2.80)$$

where the integration range is the domain of definition of the variable x . Introducing the notation $T(y, x) \equiv W(x + y|x)$, the master equation (2.80) can be rewritten as

$$\frac{\partial p}{\partial t}(x, t) = \int dy [T(y|x - y)p(x - y, t) - T(y, x)p(x, t)]. \quad (2.81)$$

It is then possible to expand the dependence on $x - y$ of $T(y|x - y)p(x - y, t)$ in powers of y around $x - y = x$, leading to the Taylor series expansion

$$T(y|x - y)p(x - y, t) = \sum_{n=0}^{\infty} \frac{(-y)^n}{n!} \frac{\partial^n}{\partial x^n} [T(y, x)p(x, t)] \quad (2.82)$$

with the convention that the zeroth order derivative of a function is the function itself. Reporting expansion (2.82) into Eq. (2.81) and exchanging the order of integrals and derivatives, we get the so-called Kramers-Moyal expansion [1, 4] of the master equation,

$$\frac{\partial p}{\partial t}(x, t) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} [\alpha_n(x)p(x, t)] \quad (2.83)$$

where we have defined

$$\alpha_n(x) \equiv \int dy y^n W(x + y|x). \quad (2.84)$$

Truncating this expansion to second order in the derivatives, one obtains the Fokker-Planck equation

$$\frac{\partial p}{\partial t}(x, t) = -\frac{\partial}{\partial x} [\alpha_1(x)p(x, t)] + \frac{\partial^2}{\partial x^2} [\alpha_2(x)p(x, t)]. \quad (2.85)$$

The coefficients $\alpha_1(x)$ and $\alpha_2(x)$ play the same role as the quantities $Q(x)$ and $\Gamma/2$ appearing in Eq. (2.77). Note that the coefficient $\alpha_2(x)$ is here a function of x , while its counterpart Γ was assumed to be constant in Eq. (2.77).

2.3.3 More General Forms of the Fokker-Planck Equation

The Fokker-Planck equation can also be generalized to an arbitrary number of coupled variables. Let us consider a set of N variables $x_i(t)$ obeying the Langevin dynamics

$$\frac{dx_i}{dt} = Q_i(x_1, \dots, x_N) + \xi_i(t), \quad i = 1, \dots, N, \quad (2.86)$$

where the N random noises $\xi_i(t)$ are correlated according to

$$\langle \xi_i(t) \xi_j(t') \rangle = \Gamma_{ij} \delta(t - t'). \quad (2.87)$$

The associated Fokker-Planck equation ruling the probability distribution $P(x_1, \dots, x_N, t)$ reads

$$\frac{\partial P}{\partial t} = - \sum_{i=1}^N \frac{\partial}{\partial x_i} (Q_i P) + \frac{1}{2} \sum_{i,j=1}^N \Gamma_{ij} \frac{\partial^2 P}{\partial x_i \partial x_j} \quad (2.88)$$

where we have dropped the arguments of the functions Q_i and P to lighten notations.

Finally, the Fokker-Planck equation can also be generalized to the case of multiplicative noise in the associated Langevin equation. For simplicity, we present here only the case of a single variable, but the interested reader may find more details, e.g., in Ref. [4]. We start from the multiplicative Langevin equation (2.52), that we rewrite here for clarity,

$$\frac{dx}{dt} = Q(x) + B(x) \xi(t). \quad (2.89)$$

We recall that the white noise $\xi(t)$ has a unit amplitude, and satisfies $\langle \xi(t) \xi(t') \rangle = \delta(t - t')$. As we have seen in Sect. 2.2.3, such a multiplicative Langevin equation may be interpreted in different ways, and the interpretation scheme considered has to be specified. In the Ito interpretation, the associated Fokker-Planck equation is given by

$$\frac{\partial p}{\partial t}(x, t) = - \frac{\partial}{\partial x} (Q(x) p(x, t)) + \frac{1}{2} \frac{\partial}{\partial x} \left(B(x)^2 \frac{\partial p}{\partial x} \right), \quad (2.90)$$

while in the Stratonovich interpretation, the Fokker-Planck equation reads

$$\frac{\partial p}{\partial t}(x, t) = - \frac{\partial}{\partial x} \left(Q(x) p(x, t) \right) + \frac{1}{2} \frac{\partial}{\partial x} \left(B(x) \frac{\partial}{\partial x} \left(B(x) p(x, t) \right) \right). \quad (2.91)$$

As discussed in Sect. 2.2.3, the choice of the interpretation framework depends on the problem considered. For systems with small, but non-zero correlation noise, the Stratonovich interpretation is the relevant one, hence its widespread use in physics for instance.

2.4 Anomalous Diffusion: Scaling Arguments

In the previous sections, we have introduced basic notions on random walks and some related types of stochastic processes. An important feature of standard random walks is that transitions between sites are characterized by well-defined time and length scales: the length scale is the lattice spacing a , and the time scale is the time step Δt for discrete time dynamics, or the inverse of the frequency ν for continuous time dynamics—see Sect. 2.1.3. One can also consider random walks with a continuous distribution of jump length Δx , instead of a random walk on a lattice where Δx is constrained to be an integer multiple of the lattice spacing a . Yet, as long as the second moment $\langle (\Delta x)^2 \rangle$ of the jump length remains finite, the basic properties of the random walk remain the same. In particular, for an unbiased random walk such that $\langle \Delta x \rangle = 0$, the mean-square displacement $\langle x(t)^2 \rangle$ is proportional to t independently of the detailed shape of the distribution $P(\Delta x)$. In other words, the typical displacement $x_{\text{typ}}(t)$ is of the order of $t^{1/2}$.

However, if the (symmetric) distribution $P(\Delta x)$ is broad enough so that $\langle (\Delta x)^2 \rangle$ is infinite, the typical displacement $x_{\text{typ}}(t)$ generically grows faster than $t^{1/2}$, often as a power t^β with $\beta > 1/2$. Such a random walk is called superdiffusive, as it moves faster than a diffusive walk. This happens in particular when $P(\Delta x)$ has power-law tails

$$P(\Delta x) \sim \frac{1}{|\Delta x|^{1+\alpha}}, \quad |\Delta x| \rightarrow \infty \quad (2.92)$$

with $0 < \alpha < 2$ (the symbol \sim here means asymptotic proportionality). Similarly, instead of considering broadly distributed jumps, one may consider a broad distribution of the time τ elapsed between two successive jumps, leading to an intermittent dynamics. Such random walks are generically called continuous time random walks. If the distribution $\psi(\tau)$ is such that its first moment $\langle \tau \rangle$ is infinite (which typically happens when $\psi(\tau) \sim 1/\tau^{1+\alpha}$ with $0 < \alpha < 1$), the dynamics is slowed down, and the typical displacement grows more slowly than $t^{1/2}$, often as a power t^β with $\beta < 1/2$. Such a random walk is called subdiffusive.

A well-defined mathematical formalism exists to properly deal with such continuous time random walks—see, e.g., [5, 6]. Some aspects of these anomalous random walks can also be studied in the framework of the Generalized Central Limit Theorem, which is introduced in Chap. 6. Here, we simply wish to provide the reader with some simple scaling arguments that can be used to understand some of the basic properties of anomalous random walks.

2.4.1 Importance of the Largest Events

Qualitatively, the reason why anomalous random walks have a typical displacement that scales differently from $t^{1/2}$, is that extreme events (very large jumps, or very large time lags between two jumps) start to play an important role. This is actually a

major property of broad distributions. In this subsection, we would like first to give a flavour, in intuitive terms, of why these large events acquire a significant statistical weight. To do so, we consider a positive random variable x with a distribution $p(x)$ having a power-law tail

$$p(x) \sim \frac{1}{x^{1+\alpha}}, \quad x \rightarrow \infty. \quad (2.93)$$

When α is lowered, the distribution (2.93) becomes broader and broader, with a ‘heavy tail’ that contains a significant part of the probability weight. In other words, very large values of x have a significant probability to be drawn from the distribution, and such large values play an essential role in the sum.

We focus on the regime where this effect is the strongest, which corresponds to $\alpha < 1$. Indeed, in this range of α , the average value $\langle x \rangle$ itself becomes infinite. Considering N random values x_i , $i = 1, \dots, N$ drawn from the distribution $p(x)$, we wish to compare the largest value in the set $\{x_i\}$ to the sum $\sum_{i=1}^N x_i$. The typical value of the maximum $\max(x_i)$ can be evaluated as follows. Let us define

$$F_N^{\max}(z) \equiv \text{Prob}\left(\max(x_1, \dots, x_N) < z\right). \quad (2.94)$$

From the independence property of the x_i ’s, one has

$$F_N^{\max}(z) = \left(\int_{-\infty}^z p(x) dx\right)^N = \left(1 - \tilde{F}(z)\right)^N, \quad (2.95)$$

where we have defined the complementary cumulative distribution $\tilde{F}(z) \equiv \int_z^{\infty} p(x) dx$. As the typical value of $\max(x_1, \dots, x_N)$ is large for large N , we can approximate $\tilde{F}(z)$ by its asymptotic behavior at large z :

$$\tilde{F}(z) \approx \frac{c}{z^\alpha}, \quad z \rightarrow \infty \quad (2.96)$$

where we have introduced explicitly the proportionality constant c . It follows that

$$\ln\left(1 - \tilde{F}(z)\right)^N \approx -\frac{cN}{z^\alpha} \quad (2.97)$$

so that

$$F_N^{\max}(z) \approx e^{-cN/z^\alpha}. \quad (2.98)$$

In other words, $F_N^{\max}(z)$ can be rewritten in the scaling form

$$F_N^{\max}(z) \approx \Phi\left(\frac{z}{N^{1/\alpha}}\right), \quad (2.99)$$

with $\Phi(u) = e^{-cu^{-\alpha}}$, which indicates that the typical value of $\max(x_i)$ is of the order of $N^{1/\alpha}$, as $F_N^{\max}(z)$ increases from 0 to 1 around $z \approx N^{1/\alpha}$. Note that Eq. (2.99) is precisely a definition of the notion of typical value: it is the value by which the variable needs to be rescaled in the expression of the probability distribution (either the cumulative probability distribution, or the probability density).

The observation that the typical value of z is of the order of $N^{1/\alpha}$ has important consequences on the sum $\sum_{i=1}^N x_i$. Intuitively, one expects the typical value of the sum to be proportional to the number N of terms. If $\alpha > 1$, $N^{1/\alpha} \ll N$ for large N , so that the largest term remains much smaller than the sum. In contrast, if $\alpha < 1$, $N^{1/\alpha} \gg N$, and the assumption that $\sum_{i=1}^N x_i$ is of the order of N breaks down, as the sum is necessarily greater than its largest term (we recall that all terms are positive). A more involved study shows in this case that the sum is of the order of the largest term itself, namely

$$\sum_{i=1}^N x_i \sim N^{1/\alpha}. \quad (2.100)$$

It is then customary to say that the largest term ‘dominates’ the sum.

For $1 < \alpha < 2$, the situation is slightly more subtle: the largest term remains much smaller than the sum, consistently with the finiteness of $\langle x \rangle$ which implies $\sum_{i=1}^N x_i \sim N \langle x \rangle$. However, the fluctuations of x remain large, as witnessed by the divergence of the variance of x , which prevents the Central Limit Theorem for being applicable—see Chap. 6.

2.4.2 Superdiffusive Random Walks

The above behavior of the statistics of a sum of broadly distributed random variables has important consequences for anomalous diffusion processes. Let us start with the superdiffusive case, which corresponds to a broad distribution of jump sizes. We consider a discrete time random walk evolving according to $x_{t+1} = x_t + u_t$, where u_t is drawn from a symmetric distribution $p(u)$. We assume that space is continuous, and that the variables $\{u_t\}$ are independent and identically distributed random variables. Accordingly, the position x_t is given by

$$x_t = \sum_{t'=0}^{t-1} u_{t'} \quad (2.101)$$

where we have assumed that $x_0 = 0$. The present problem is thus directly related to problems of random sums. The symmetry of the distribution $p(u)$ implies that $\langle u_t \rangle = 0$, from which $\langle x_t \rangle = 0$ follows. If $\langle u^2 \rangle$ is finite, one has

$$\langle x_t^2 \rangle = \sum_{t', t''} \langle u_{t'} u_{t''} \rangle = t \langle u^2 \rangle \quad (2.102)$$

where we have used the fact that the variables $u_{t'}$ and $u_{t''}$ are statistically independent for $t' \neq t''$, implying $\langle u_{t'} u_{t''} \rangle = 0$. Hence the mean-square displacement $\langle x_t^2 \rangle$ is linear in t , which corresponds to a normal diffusive process.

In contrast, if the distribution $p(u)$ is broad, with an infinite variance, the above reasoning fails, since the average values appearing in Eq. (2.102) are infinite. Let us consider for definiteness a distribution $p(u)$ such that

$$p(u) \sim \frac{1}{|u|^{1+\alpha}}, \quad u \rightarrow \pm\infty, \quad (2.103)$$

with $\alpha < 2$, so that $\langle u^2 \rangle$ is infinite—see Fig. 2.2 for an illustration. We can however use a scaling argument inspired by Eq. (2.102), by using typical values instead of average values:

$$x_{\text{typ}}(t)^2 \sim \sum_{t'=0}^{t-1} u_{t'}^2 \quad (2.104)$$

where we have neglected cross terms $u_{t'} u_{t''}$ ($t' \neq t''$). The typical value of the square displacement is thus the typical value of a sum of t independent random variables $y_{t'} \equiv u_{t'}^2$. Using the relation

$$P(y) = p(u) \left| \frac{du}{dy} \right|, \quad (2.105)$$

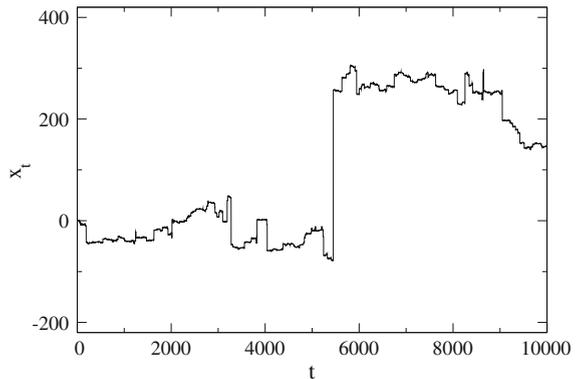
one finds that the distribution $P(y)$ also has a power-law tail, satisfying

$$P(y) \sim \frac{1}{y^{1+\frac{\alpha}{2}}}, \quad y \rightarrow \infty. \quad (2.106)$$

Combining Eqs. (2.100) and (2.104), one then deduces that

$$x_{\text{typ}}(t) \sim t^{1/\alpha}. \quad (2.107)$$

Fig. 2.2 Illustration of a superdiffusive random walk, with a power-law distribution of jump sizes, of parameter $\alpha = 0.8$ —see Eq. (2.103). The largest jump is of the same order as the total displacement, whatever the chosen time window



Hence the random walk has an anomalous behavior of its typical displacement, $x_{\text{typ}}(t) \sim t^\beta$, characterized by an exponent $\beta = 1/\alpha > 1/2$ (since $\alpha < 2$). The walk is thus superdiffusive. Note that a more rigorous derivation of this result can be carried out using the Generalized Central Limit Theorem, introduced in Chap. 6.

2.4.3 Subdiffusive Random Walks

On the contrary, subdiffusive walks have (in the simplest cases) a well-defined jump length, but exhibit strong local trapping effects, so that the sojourn times on a given site become broadly distributed, instead of being fixed to a value Δt as in the above superdiffusive example. We thus consider a random walk process in which the time lag τ between two jumps is itself a random variable τ following a distribution $p(\tau)$, with a tail

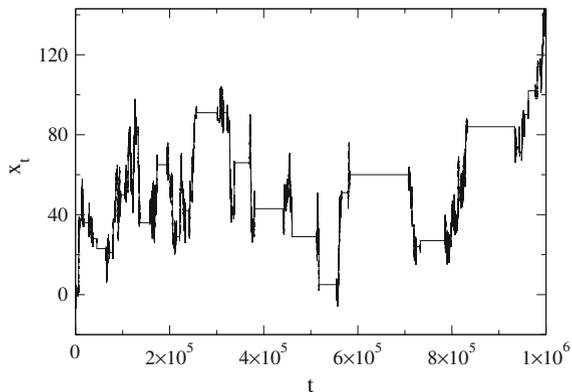
$$p(\tau) \sim 1/\tau^{1+\alpha}, \quad \tau \rightarrow \infty \quad (0 < \alpha < 1). \quad (2.108)$$

After a time τ , the walker jumps to one of the two neighboring sites, namely $x_{t+\tau} = x_t + \epsilon_t$, where $\epsilon_t = \pm 1$ with equal probabilities. An illustration is provided in Fig. 2.3.

Here again, the behavior of the random walk can be understood through a simple scaling argument. After N steps, the typical displacement x_N^{typ} of the walker is of the order of \sqrt{N} . To relate N to the actual time t , one can observe that time t is the sum of the N sojourn times τ_i at the i th position. Hence, using the estimation given in Eq. (2.100), one finds

$$t = \sum_{i=1}^N \tau_i \sim N^{1/\alpha} \quad (2.109)$$

Fig. 2.3 Illustration of a subdiffusive random walk, with a power-law distribution of sojourn times of parameter $\alpha = 0.8$, as defined in Eq. (2.108). The largest sojourn time is typically a finite fraction of the total time window



whence the scaling $N \sim t^\alpha$ follows. Combining this relation with $x_N^{\text{typ}} \sim N^{1/2}$, we finally obtain

$$x_{\text{typ}}(t) \sim t^{\alpha/2}. \quad (2.110)$$

The exponent β characterizing the anomalous diffusion is thus $\beta = \alpha/2 < 1/2$ (we recall that $\alpha < 1$). The random walk is therefore slower than normal diffusion, or in other words, subdiffusive. Note that this case is not a direct application of the Generalized Central Limit Theorem, but there however exist rigorous methods to derive the above scaling behaviors.

To conclude this section on anomalous diffusion, it is interesting to mention a related situation which leads to a different scaling exponent. In the above subdiffusive walk, the lag time τ between two successive jumps is drawn anew at each step, and is independent of all the previous values of τ . However, thinking of the physical situation of a particle randomly evolving in a complex (though one-dimensional) potential energy landscape, one may identify the local minima of the potential with the sites of a lattice. Jumps from one local minimum to a neighboring one occur through thermal activation over the energy barrier separating the two minima. The sojourn time in a minimum, given by an Arrhenius law, is exponential with respect to the energy barrier, and is thus very sensitive to the value of the barrier. The presence of an even relatively moderate range of values of the energy barriers may lead in some regime to broad distributions of sojourn times. However, the main difference with the case studied above is that the sojourn time is approximately the same at each visit on the same site. This situation can be qualified as a ‘frozen disorder’. The successive time lags are thus no longer statistically independent, as was the case above. Hence the scaling argument needs to be modified accordingly. We still have that the typical displacement after N steps is $x_N^{\text{typ}} \sim \sqrt{N}$. This means that the number of distinct sites visited by the walk during N steps is also of the order of \sqrt{N} . Each of these sites has been visited of the order of $N/\sqrt{N} = \sqrt{N}$ times. Hence the total time t elapsed after N steps can be roughly approximated as

$$t \sim \sqrt{N} \sum_{i=1}^{\sqrt{N}} \tau_i \quad (2.111)$$

(the upper bound in the sum should be understood as the integer part of \sqrt{N}). The variables τ_i under the sum correspond to the (fixed) sojourn times in the \sqrt{N} distinct visited sites, while the factor \sqrt{N} in front of the sum is the typical number of visits per site. Altogether, we have if $p(\tau) \sim 1/\tau^{1+\alpha}$ when $\tau \rightarrow \infty$ (with $\alpha < 1$) that

$$t \sim \sqrt{N}^{1+\frac{1}{\alpha}}. \quad (2.112)$$

It follows that $\sqrt{N} \sim t^{\alpha/(1+\alpha)}$, so that we finally obtain

$$x_{\text{typ}}(t) \sim t^\beta, \quad \beta = \frac{\alpha}{1+\alpha} < \frac{1}{2}, \quad (2.113)$$

resulting again in a subdiffusive motion, but with an exponent β different from the exponent $\alpha/2$ obtained in Eq. (2.110). Interestingly, one has

$$\frac{\alpha}{1 + \alpha} > \frac{\alpha}{2} \quad (2.114)$$

so that the presence of fixed sojourn times on each site (frozen disorder) actually leads to a faster motion than the ‘annealed’ case where the sojourn time is drawn anew at each visit. This can be understood from the fact that only \sqrt{N} random variables τ_i are drawn in the frozen disorder case, while N sojourn times are picked up in the ‘annealed’ case. Drawing more random variables typically leads to the exploration of longer sojourn times τ , which slows down the dynamics.

2.5 Fast and Slow Relaxation to Equilibrium

2.5.1 Relaxation to Canonical Equilibrium

Up to now, we have mostly considered steady-state statistical properties, although we already mentioned time dependent situations when considering random walks. Here, we wish to explicitly discuss the convergence of the probability distribution of configurations to the equilibrium distribution. To this aim, let us consider a stochastic process with n energy states E_i , $i = 1, \dots, n$. The continuous time stochastic process is defined by transition rates W_{ji} from configuration i with energy E_i , to configuration j with energy E_j ($i \neq j$)—we use here matrix notations for the transition rates for reasons that will become clear below. These transition rates are assumed to obey the following detailed balance relation,

$$W_{ji} e^{-\beta E_i} = W_{ij} e^{-\beta E_j} \quad (2.115)$$

for all pairs (i, j) (with β the inverse temperature), so that the equilibrium probability distribution reads

$$P_i^{\text{eq}} = \frac{1}{Z} e^{-\beta E_i}, \quad Z = \sum_{j=1}^n e^{-\beta E_j}. \quad (2.116)$$

The master equation Eq. (2.9) governing the probability $P_i(t)$ to be in configuration i at time t reads with the current notations

$$\frac{dP_i}{dt} = \sum_{j(j \neq i)} \left(W_{ij} P_j(t) - W_{ji} P_i(t) \right). \quad (2.117)$$

Defining the coefficient W_{ii} as

$$W_{ii} = - \sum_{j(j \neq i)} W_{ji} \quad (2.118)$$

we can rewrite the master equation as

$$\frac{dP_i}{dt} = \sum_{j=1}^n W_{ij} P_j(t) \quad (2.119)$$

which shows that this equation takes the form of a matrix equation. We may thus rewrite the equation more formally as

$$\frac{d}{dt} \mathbf{P}(t) = \mathbf{W} \mathbf{P}(t) \quad (2.120)$$

where $\mathbf{P}(t)$ is the vector of components $(P_1(t), \dots, P_n(t))$ and \mathbf{W} is the matrix of elements (W_{ij}) , $i, j = 1, \dots, n$. In this form, finding the stationary distribution amounts to an eigenvalue problem, namely finding the eigenvector \mathbf{P}^{eq} associated to the eigenvalue $\lambda = 0$ of the matrix \mathbf{W} . However, this matrix reformulation potentially provides more information than just the stationary distribution. The other, nonzero, eigenvalues (that can be shown to be negative) precisely describe the relaxation of the distribution to the equilibrium value. Diagonalizing the matrix \mathbf{W} , one eventually obtains

$$\mathbf{P}(t) = \sum_{j=1}^n e^{\lambda_j t} \mathbf{Q}^{(j)} \quad (2.121)$$

where the λ_j 's are the eigenvalues of the matrix \mathbf{W} (labelled in decreasing order such that $\lambda_1 = 0$ and $\lambda_n < \dots < \lambda_2 < 0$), and the $\mathbf{Q}^{(j)}$'s are vectors depending on the eigenvectors of \mathbf{W} and on the initial distribution $\mathbf{P}(t = 0)$. It is clear from Eq. (2.121) that $\mathbf{P}(t)$ converges to $\mathbf{Q}^{(1)}$ when $t \rightarrow \infty$, so that $\mathbf{Q}^{(1)}$ identifies with the equilibrium distribution \mathbf{P}_{eq} . Note that the normalization condition $\sum_{i=1}^n P_i(t) = 1$ implies that, for $j > 1$,

$$\sum_{i=1}^n Q_i^{(j)} = 0. \quad (2.122)$$

At large enough times, only the most slowly decreasing term in Eq. (2.121) contributes on top of the equilibrium distribution, and the probability distribution can be approximated as

$$\mathbf{P}(t) \approx \mathbf{P}_{\text{eq}} + e^{-\lambda_2 t} \mathbf{Q}^{(2)}. \quad (2.123)$$

These generic properties can easily be illustrated in the case of a two-state system ($n = 2$). Given the normalization constraint $P_1(t) + P_2(t) = 1$, the evolution of the probabilities can be expressed only in terms of $P_1(t)$, leading to

$$\frac{dP_1}{dt} = -\left(W_{12} + W_{21}\right)P_1(t) + W_{12} \quad (2.124)$$

whose solution is readily obtained as

$$P_1(t) = P_1^{\text{eq}} + e^{-\alpha t}(P_1(0) - P_1^{\text{eq}}) \quad (2.125)$$

with $\alpha = W_{12} + W_{21}$ and, assuming that the detailed balance relation (2.115) holds

$$P_1^{\text{eq}} = \frac{W_{12}}{W_{12} + W_{21}} = \frac{e^{-\beta E_1}}{e^{-\beta E_1} + e^{-\beta E_2}}. \quad (2.126)$$

In this simple case, the relaxation to equilibrium is purely exponential at all times, and not only asymptotically at large times.

2.5.2 Dynamical Increase of the Entropy

Another way to characterize the relaxation to equilibrium is to show that relaxation is accompanied by an increase of entropy, in agreement with the second law of thermodynamics. To this aim, we first need to introduce a time-dependent entropy defined as

$$S(t) = - \sum_C P(C, t) \ln P(C, t). \quad (2.127)$$

This definition closely follows the definition (1.71). Coming back to usual notations for the transition rates, we assume that $W(C'|C) = W(C|C')$, which is a specific form of the detailed balance relation (2.12) associated to a uniform equilibrium distribution over a subset of configurations (as in the microcanonical ensemble). Under this assumption, one can show that $S(t)$ is an increasing function of time. Let us start by computing the time-derivative of the entropy:

$$\frac{dS}{dt} = - \sum_C \frac{dP}{dt}(C, t) \ln P(C, t) - \sum_C \frac{dP}{dt}(C, t). \quad (2.128)$$

The last term cancels out due to the normalization condition $\sum_C P(C, t) = 1$. Using the master equation, one has:

$$\begin{aligned} \frac{dS}{dt} &= - \sum_C \ln P(C, t) \sum_{C'(\neq C)} (-W(C'|C)P(C, t) + W(C|C')P(C', t)) \\ &= \sum_{C, C'(C \neq C')} \ln P(C, t) (W(C'|C)P(C, t) - W(C|C')P(C', t)). \end{aligned} \quad (2.129)$$

Exchanging the notations C and C' in the last equation, we also have

$$\frac{dS}{dt} = \sum_{C, C'(C \neq C')} \ln P(C', t) (W(C|C')P(C', t) - W(C'|C)P(C, t)). \quad (2.130)$$

Summing Eqs. (2.129) and (2.130), and using the detailed balance property $W(C'|C) = W(C|C')$, we obtain

$$\frac{dS}{dt} = \frac{1}{2} \sum_{C, C'(C \neq C')} (\ln P(C', t) - \ln P(C, t)) (P(C', t) - P(C, t)) W(C|C'). \quad (2.131)$$

As $[\ln P(C', t) - \ln P(C, t)]$ and $[P(C', t) - P(C, t)]$ have the same sign, one concludes that

$$\frac{dS}{dt} \geq 0. \quad (2.132)$$

This is one possible statement, in the context of stochastic processes, of the second law of thermodynamics. Moreover, in the stationary state, $dS/dt = 0$, and one necessarily has for all pairs (C, C') either $P_{\text{st}}(C) = P_{\text{st}}(C')$ or $W(C|C') = 0$, where $P_{\text{st}}(C)$ is the stationary probability distribution. One then recovers, consistently with the detailed balance assumption $W(C'|C) = W(C|C')$, the postulate of equilibrium statistical mechanics stating that mutually accessible configurations have the same probability.

More generally, for Markovian stochastic processes described by the master equation (2.9), it is always possible to define a functional $\tilde{S}(\{P(C, t)\})$ that increases with time, without need for detailed balance or microreversibility properties [1]. The general definition of $\tilde{S}(\{P(C, t)\})$ is

$$\tilde{S}(t) = - \sum_C P(C, t) \ln \left(\frac{P(C, t)}{P_{\text{st}}(C)} \right). \quad (2.133)$$

A drawback of this definition is that the stationary distribution $P_{\text{st}}(C)$ needs to be known in order to define \tilde{S} , which in many cases restricts the usefulness of the functional \tilde{S} . Yet, a simple application of the generalized definition (2.133) of the entropy is the case of the canonical ensemble. Using $P_{\text{st}} \propto e^{-\beta E}$, one finds

$$\tilde{S}(t) = \beta F_{\text{eq}} + S(t) - \beta \langle E(t) \rangle \quad (2.134)$$

where F_{eq} is the (equilibrium) free energy, and $S(t)$ the time-dependent entropy defined in Eq. (2.127). This definition suggests to introduce a time-dependent free energy $F(t)$ as

$$F(t) = \langle E(t) \rangle - TS(t) \quad (2.135)$$

with $T = \beta^{-1}$ the temperature. In this way, one has

$$\tilde{S}(t) = \beta(F_{\text{eq}} - F(t)), \quad (2.136)$$

leading to $dF/dt \leq 0$. Hence the time-dependent free-energy is a decreasing function of time. This result is consistent with the standard thermodynamic result that the free energy of a system in contact with a thermostat can only decrease under spontaneous evolution.

2.5.3 *Slow Relaxation and Physical Aging*

Although many systems converge to a stationary state on times shorter than or comparable to the observation time, it turns out that some systems do not reach a steady state and keep evolving on time scales that can be very large compared to standard observation times. This is the case for instance of glasses, which keep aging for years or more, as well as in laser cooling experiments [7]. It is also likely that aging mechanisms, or slow relaxation effects, play a significant role in many different types of complex systems. Even though the aging mechanisms may differ from one situation to the other, it is certainly of interest to investigate one of the simplest known aging phenomena, illustrated by the trap model, which we describe here within a generic formalism that does not rely on a specific physical realization.

Let us consider a model system in which to each configuration C is associated a given lifetime τ . This lifetime τ is the mean time spent in configuration C before moving to another configuration. As we consider only temporal aspects of the dynamics, and not other types of observables (energy, magnetization,...), we simply label the configurations by their lifetime τ . We then choose a simple form for the transition rate $W(\tau'|\tau)$, namely:

$$W(\tau'|\tau) = \frac{1}{\tau}\psi(\tau'). \quad (2.137)$$

The function $\psi(\tau')$ is the a priori probability distribution of the configurations τ' , meaning that the selected new configuration is chosen completely at random. From the normalization condition $\int_0^\infty d\tau'\psi(\tau') = 1$, we have

$$\int_0^\infty d\tau' W(\tau'|\tau) = \frac{1}{\tau}, \quad (2.138)$$

so that the characteristic escape time from a configuration with lifetime τ is precisely τ , as it should. For simplicity, we also assume that all lifetimes τ are greater than a value τ_0 , that we set to $\tau_0 = 1$ in the following. The master equation then reads:

$$\begin{aligned}
\frac{\partial P}{\partial t}(\tau, t) &= -P(\tau, t) \int_1^\infty d\tau' W(\tau'|\tau) + \int_1^\infty d\tau' W(\tau|\tau') P(\tau', t) \\
&= -\frac{1}{\tau} P(\tau, t) + \psi(\tau) \int_1^\infty \frac{d\tau'}{\tau'} P(\tau', t).
\end{aligned} \tag{2.139}$$

At equilibrium, the probability to be in a configuration with lifetime τ is proportional to τ and to the a priori distribution $\psi(\tau)$ of configurations:

$$P_{\text{eq}}(\tau) = \frac{1}{\langle \tau \rangle} \tau \psi(\tau), \tag{2.140}$$

where $\langle \tau \rangle$ is defined as

$$\langle \tau \rangle = \int_1^\infty d\tau \tau \psi(\tau). \tag{2.141}$$

Similarly to the case of anomalous diffusion discussed in Sect. 2.4, the key ingredient that determines the behavior of the process is the shape of the tail of the lifetime distribution $\psi(\tau)$. The most interesting situation corresponds to a distribution $\psi(\tau)$ with a power-law tail. Here, for simplicity, we take a distribution with a pure power-law form, namely

$$\psi(\tau) = \frac{\alpha}{\tau^{1+\alpha}}, \quad \tau > 1 \quad (\alpha > 0). \tag{2.142}$$

An example of physical realization is the case of a particle trapped into potential wells of random depth E , with an exponential distribution

$$\rho(E) = \frac{1}{E_0} e^{-E/E_0}. \tag{2.143}$$

The lifetime τ is given by the standard Arrhenius law

$$\tau = \tau_0 e^{E/T}, \tag{2.144}$$

where $\tau_0 = 1$ is a microscopic time scale. Using the relation $\psi(\tau)|d\tau| = \rho(E)|dE|$, one precisely finds the form (2.142) for $\psi(\tau)$, with $\alpha = T/E_0$.

In the case $\alpha > 1$, $\langle \tau \rangle$ is finite, but if $\alpha \leq 1$ then $\langle \tau \rangle$ is infinite, so that the equilibrium distribution (2.140) does not exist, as it is not normalizable. As a result, no stationary state can be reached, and the system keeps drifting towards configurations with larger and larger lifetimes τ .

It is then of interest to determine the time-dependent probability distribution $P(\tau, t)$ in the long-time regime. We postulate the following scaling form

$$P(\tau, t) = \frac{1}{t} \phi\left(\frac{\tau}{t}\right). \tag{2.145}$$

From the normalization condition of $P(\tau, t)$, one has

$$\int_1^\infty d\tau P(\tau, t) = \frac{1}{t} \int_1^\infty d\tau \phi\left(\frac{\tau}{t}\right) = 1, \quad (2.146)$$

from which one gets, with the change of variable $u = \tau/t$,

$$\int_{1/t}^\infty du \phi(u) = 1. \quad (2.147)$$

As $\phi(u)$ does not depend explicitly on time t , the above condition cannot be satisfied for all t . But we are looking for an asymptotic large- t solution, so that we impose that Eq. (2.147) is satisfied in the infinite t limit, namely

$$\int_0^\infty du \phi(u) = 1. \quad (2.148)$$

As a result, the scaling form (2.145) is an approximate solution that becomes exact when $t \rightarrow \infty$. Equation (2.145) yields for the time derivative of $P(\tau, t)$:

$$\frac{\partial P}{\partial t} = -\frac{1}{t^2} \phi\left(\frac{\tau}{t}\right) - \frac{\tau}{t^3} \phi'\left(\frac{\tau}{t}\right), \quad (2.149)$$

where ϕ' is the derivative of ϕ . Multiplying Eq. (2.149) by t^2 , one obtains, with the notations $u = \tau/t$ and $v = \tau'/t$,

$$-\phi(u) - u\phi'(u) = -\frac{1}{u}\phi(u) + \psi(ut) t \int_{1/t}^\infty \frac{dv}{v} \phi(v). \quad (2.150)$$

Using the specific form (2.142) of $\psi(\tau)$, we find

$$\left(1 - \frac{1}{u}\right) \phi(u) + u\phi'(u) + \frac{\alpha}{u^{1+\alpha}} t^{-\alpha} \int_{1/t}^\infty \frac{dv}{v} \phi(v) = 0. \quad (2.151)$$

For the above equation to be well-defined in the infinite t limit in which it is supposed to be valid, the explicit t -dependence has to cancel out. One thus needs to have

$$\int_{1/t}^\infty \frac{dv}{v} \phi(v) \sim t^\alpha, \quad t \rightarrow \infty, \quad (2.152)$$

which requires that $\phi(v)$ has the following asymptotic form at small v :

$$\phi(v) \approx \frac{\phi_0}{v^\alpha}, \quad v \rightarrow 0. \quad (2.153)$$

Here, ϕ_0 is an unknown constant, to be determined later on from the normalization condition of $\phi(u)$. The master equation is then finally written as the following differential equation:

$$\left(1 - \frac{1}{u}\right) \phi(u) + u\phi'(u) + \frac{\phi_0}{u^{1+\alpha}} = 0. \quad (2.154)$$

This equation is a linear inhomogeneous differential equation, and its solution can be found using standard techniques. The solution of Eq. (2.154) satisfying the normalization condition (2.148) reads [8]

$$\phi(u) = \frac{\sin(\pi\alpha)}{\Gamma(\alpha)} \frac{1}{u} e^{-1/u} \int_0^{1/u} dv v^{\alpha-1} e^v, \quad (2.155)$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is the Euler Gamma function. It is rather easy to show that $\phi(u) \sim u^{-\alpha}$ for $u \rightarrow 0$ as expected, and that $\phi(u) \sim u^{-1-\alpha}$ for $u \rightarrow \infty$, leading for $P(\tau, t)$ to

$$P(\tau, t) \propto \tau\psi(\tau), \quad \tau \ll t, \quad (2.156)$$

$$P(\tau, t) \propto \psi(\tau), \quad \tau \gg t. \quad (2.157)$$

These asymptotic behaviors can be interpreted rather easily: configurations with lifetimes $\tau \ll t$ have been visited a large number of times, so that they are quasi-equilibrated; in contrast, configurations with lifetimes $\tau \gg t$ have been visited at most once, and the precise value of τ is not yet felt by the dynamics.

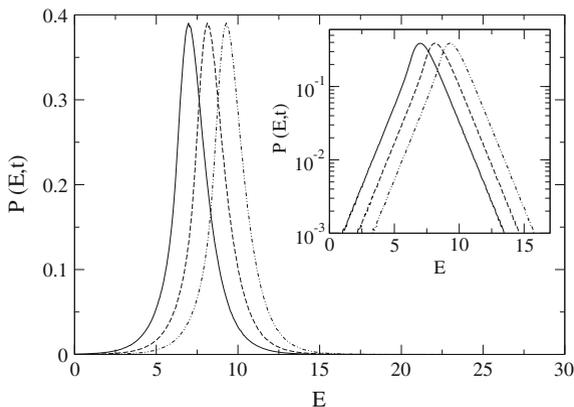


Fig. 2.4 Energy barrier distribution in the aging regime of the trap model ($T = T_g/2$), for different times $t = 10^6$ (full line), $t = 10^7$ (dashed line) and $t = 10^8$ (dot-dashed). The distribution drifts toward larger energy barriers, logarithmically with time. The inset shows the same data on a semi-logarithmic scale, to visualize the exponential tails

In the physical example of the trap model defined by Eqs. (2.143) and (2.144), the aging regime occurs for $\alpha = T/E_0 < 1$, so that $T_g \equiv E_0$ turns out to be the glass transition temperature. For $T < T_g$, the energy distribution $p(E, t)$, obtained from $P(\tau, t)$ by a simple change of variable, takes the scaling form $p(E, t) = \Phi(E - T \ln t)$. A logarithmic drift toward larger energy barriers is thus observed, as illustrated in Fig. 2.4. The average energy is given by

$$\bar{E}(t) \approx E_1 + T \ln t \quad (2.158)$$

where E_1 is a temperature-dependent constant.

References

1. N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, 1992)
2. B.V. Gnedenko, A.N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley, 1954)
3. V.V. Petrov, *Sums of Independent Random Variables* (Springer, Berlin, 1975)
4. C.W. Gardiner, *Stochastic Methods*, 4th edn. (Springer, Berlin, 2009)
5. J.-P. Bouchaud, A. Georges, Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Phys. Rep.* **195**, 127 (1990)
6. R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1 (2000)
7. E. Bertin, F. Bardou, From laser cooling to aging: a unified Lévy flight description. *Am. J. Phys.* **76**, 630 (2008)
8. C. Monthus, J.-P. Bouchaud, Models of traps and glass phenomenology. *J. Phys. A* **29**, 3847 (1996)

Chapter 3

Statistical Physics of Interacting Macroscopic Units

Until now, we have mainly considered physical systems, in which elementary units are implicitly atoms or molecules. In this case, the laws of motion of the individual particles are known, and the main difficulty consists in being able to change the scale of description, going from the scale of particles to the system size.

However, our everydaylife experience tells us that there exist many familiar systems that are composed of interacting macroscopic units, that thus behave very differently from atoms or molecules: examples range from sand piles, foams, bacteria colonies, animal flocks, or road traffic, to quote only a few examples. In such cases, it is clear that the interacting objects, or individuals, cannot be described in the same way as molecules, and precise dynamical laws at the individual scale are most often not known.

The difficulties encountered when trying to apply a statistical physics approach to such assemblies of macroscopic units are then two-fold. On the one-hand, a model should be given for the dynamics of individual, and it is often not clear how relevant or reliable such modeling is to describe realistic systems. On the other hand, reasonable models of individual dynamics usually do not have similar conservation laws and time-reversal symmetry as the Hamiltonian dynamics of molecular systems. Hence it is hard, even in specific cases, to build a statistical physics approach from a postulate similar to the hypothesis of equiprobability of configurations having the same energy. Interesting attempts in this direction, notably in the context of granular matter, have however been proposed [1].

In this section, we illustrate on several examples how different statistical physics techniques can be devised, in specific cases, to describe assemblies of interacting units. In the first example (the dynamics of residential moves in a city, Sect. 3.1), a mapping can be performed to an effective equilibrium system, yielding interesting insights. In the second example (Zero Range Process, Sect. 3.2), an explicit stationary solution of the master equation can be found. Finally, the last example (collective motion of active particles, Sect. 3.3) is studied through the so-called Boltzmann equation, a generic approach that can be used when interactions are limited to binary ‘collisions’, that are very localized in space and time.

3.1 Dynamics of Residential Moves

A standard example of complex system dynamics is the Schelling model which represents in a schematic way the dynamics of residential moves in a city [2, 3]. The city is modeled as a checkerboard, divided into cells. Two types of agents (say red and green) live in the city. They reside in the cells of the checkerboard, with at most one agent in each cell. Agents characterize their degree of satisfaction regarding their environment by a utility, which is a given function (the same for all agents) of the number of agents of the same type in their neighborhood. The neighborhood can be defined in different ways. One possibility would be to consider the set of nearest neighbors. However, most studies rather use the Morse neighborhood, that is the 3×3 (or sometimes 5×5) square surrounding the current cell.

Before moving, an agent chooses at random an empty cell, and evaluates the utility u_{new} associated to this new location. The agent compares this quantity to the utility u_{old} of his present location, by computing the utility difference $\Delta u = u_{\text{new}} - u_{\text{old}}$. The move is then accepted with probability $1/(1 + e^{-\Delta u/T})$. Here, T is a parameter analogous to the temperature in physical systems, that characterizes the influence of other factors, like the presence of facilities, shops, or friends, that are not explicitly taken into account in the model, but could bias the decision of moving or not. At low T , and for a large class of utility functions such that agents have a (possibly slight) preference for being with agents of the same type, a segregation phenomenon is observed when simulating the model numerically: two types of domains form, namely domains with a majority of red agents and domains with a majority of green agents. Quite surprisingly, this segregation phenomenon seems quite robust, and is also observed in the case where agents have a marked preference for mixed neighborhood.

The Schelling model in its standard form is very hard to solve analytically, and solutions are not presently known. The reason for these difficulties is mainly that the neighborhoods of two neighboring cells overlap, generating complicated correlations in the system. In order to find an analytical solution, a standard strategy is to define a variant of the model on a specific geometry that avoids these correlations. This strategy was for instance successful in the Ising model, by introducing a fully connected version of the model (see Sect. 1.4.1): assuming that all spins interact together, the phase transition could be obtained analytically in a simple way.

A straightforward application of this idea to the Schelling model a priori seems to lead to a deadlock. If an agent evaluates its utility by considering the whole city as its neighborhood, this utility will not change when moving within the city. A more interesting strategy is then to divide the city into a large number of blocks, so that agents evaluate their utility within blocks, and move from blocks to blocks. In this way, correlations between blocks may be suppressed.

3.1.1 A Simplified Version of the Schelling Model

In order to implement this strategy, we consider the following model, with a single type of agent to further simplify the derivation (the case of two different types of agents is briefly discussed below). The segregation phenomenon then corresponds to the formation of domains of different densities. The city is divided into a large number Q of blocks, each block containing H cells (a cell may be thought of as representing a flat). We assume that each cell can contain at most one agent, so that the number n_q of agents in a given block q ($q = 1, \dots, Q$) satisfies $n_q \leq H$. A microscopic configuration C of the city corresponds to the knowledge of the state (empty or occupied) of each cell. For each block q , we also introduce the density of agents $\rho_q = n_q/H$. Each agent has the same utility function $u(\rho_q)$, which describes the degree of satisfaction concerning the density of the block it is living in. The collective utility is defined as the total utility of all the agents in the city: $U(C) = \sum_q n_q u(\rho_q)$.

A dynamical rule allows the agents to move from one block to another. At each time step, one picks up at random an agent and a vacant cell, within two different blocks. The agent moves in that empty cell with probability:

$$W(C'|C) = \frac{1}{1 + e^{-\Delta u/T}}, \quad (3.1)$$

where C and C' are the configurations before and after the move respectively, and Δu is the variation of the individual utility of the chosen agent, associated to the proposed move. The parameter T has the same interpretation as in the standard Schelling model.

It is interesting at this stage to emphasize the difference between the present model and standard physical approaches. It could seem at first sight that the utility is simply the equivalent, up to a sign reversal, of the energy in physics. In the present model however, an economics perspective is adopted, so that the agents are considered as purely selfish. They make decisions only according to their own utility change Δu , and do not consider the potential impact of their decision on the other agents. In contrast, in physical models, the probability for a particle to move depends on the energy variation of the whole system, and the effect on the other particles is thus taken into account from the outset. This has important consequences, as we shall see below.

We wish to find the stationary probability distribution $P(C)$ of the microscopic configurations C . This is not an easy task in general. Yet, if we were able to show that a detailed balance relation holds in this model, we would straightforwardly get the solution. Let us assume that the individual cost Δu can be written as $\Delta u = F(C) - F(C')$, where F is a function on configuration space.¹ From Eq. (3.1), we

¹The relation $\Delta u = F(C) - F(C')$ is non-trivial, because the utility of a single agent cannot be computed from the sole knowledge of the system configuration; one also needs to know who is the considered agent, and this information is not included in the configuration C .

find that the dynamics satisfies a detailed balance relation:

$$W(C'|C)P(C) = W(C|C')P(C'), \quad (3.2)$$

with a distribution $P(C)$ given by

$$P(C) = \frac{1}{Z} e^{-F(C)/T}, \quad (3.3)$$

where Z is the analog of a partition function. It can be shown that a function F satisfying this condition is given by

$$F(C) = - \sum_q \sum_{m=0}^{n_q} u(m/H). \quad (3.4)$$

To characterize the “segregation” phenomenon, the full statistical information on the occupation number of each cell is not necessary. Instead, an aggregated description in terms of densities of the blocks turns out to be more useful. Such a coarse-grained description is obtained by aggregating all configurations with the same number of agents in each block. As there are $H!/[(n!)(H-n)!]$ ways of ordering n agents in H cells, we obtain the following coarse-grained probability distribution:

$$\tilde{P}(n_1, \dots, n_Q) = \tilde{K} \exp\left(-\frac{H}{T} \sum_q \tilde{f}(n_q)\right), \quad (3.5)$$

with \tilde{K} a normalization constant, and where we have introduced the function \tilde{f} :

$$\tilde{f}(n) = \frac{T}{H} \ln\left(\frac{n!(H-n)!}{H!}\right) - \frac{1}{H} \sum_{m=0}^{n_q} u\left(\frac{m}{H}\right). \quad (3.6)$$

The above expression suggests to consider the limit of large H in order to get a continuous formulation for \tilde{f} . Keeping constant the density of each block $\rho_q = n_q/H$ (ρ_q hence becoming a continuous variable) and expanding the factorials using Stirling's formula $\ln n! \approx n \ln n - n$, valid for large n , one obtains for $H \rightarrow \infty$

$$\frac{1}{H} \ln\left(\frac{n_q!(H-n_q)!}{H!}\right) \rightarrow \rho_q \ln \rho_q + (1 - \rho_q) \ln(1 - \rho_q). \quad (3.7)$$

Similarly, the last term in the expression of \tilde{f} converges to an integral:

$$\frac{1}{H} \sum_{m=0}^{n_q} u\left(\frac{m}{H}\right) \rightarrow \int_0^{\rho_q} u(\rho') d\rho'. \quad (3.8)$$

In terms of density ρ_q , the stationary distribution $\tilde{P}(n_1, \dots, n_Q)$ turns into a probability density $P(\rho_1, \dots, \rho_Q)$ given by (with $\sum_{q=1}^Q \rho_q = Q\rho_0$ held fixed):

$$P(\rho_1, \dots, \rho_Q) = K \exp \left(-\frac{H}{T} \sum_{q=1}^Q f(\rho_q) \right) \quad (3.9)$$

where K is a normalization constant, and where the function $f(\rho)$ is defined as

$$f(\rho) = T\rho \ln \rho + T(1 - \rho) \ln(1 - \rho) - \int_0^\rho u(\rho') d\rho'. \quad (3.10)$$

The function $\Phi(\rho_1, \dots, \rho_Q) = \sum_{q=1}^Q f(\rho_q)$ may be called a potential, or a large deviation function. It is also the analogue of the free energy functions used in physics. The configurations (ρ_1, \dots, ρ_Q) that minimize the potential $\Phi(\rho_1, \dots, \rho_Q)$ under the constraint of fixed $\sum_{q=1}^Q \rho_q$ are the most probable to come up. In the limit $H \rightarrow \infty$, these configurations are the only ones that appear in the stationary state, as the probability of other configurations vanishes exponentially with H .

3.1.2 Condition for Phase Separation

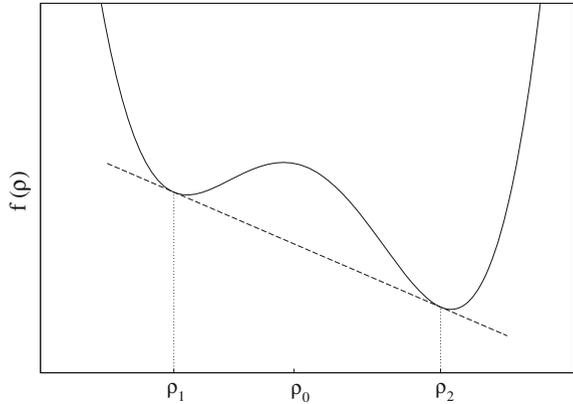
Focusing on the large H case, the problem gets back to finding the set (ρ_1, \dots, ρ_Q) which minimizes the potential $\Phi(\rho_1, \dots, \rho_Q)$ with the constraint $\sum_q \rho_q$ fixed. We are interested in knowing whether the stationary state is statistically homogeneous or inhomogeneous. Following standard physics textbooks methods [4], the homogeneous state at density ρ_0 is unstable against a phase separation if there exists two densities ρ_1 and ρ_2 such that

$$\gamma f(\rho_1) + (1 - \gamma)f(\rho_2) < f(\rho_0). \quad (3.11)$$

The parameter γ ($0 < \gamma < 1$) corresponds to the fraction of blocks that would have a density ρ_1 in the segregated state. This condition simply means that the value of the potential Φ is lower for the segregated state than for the homogeneous state, so that the segregated state has a much larger probability to occur. Geometrically, the inequality (3.11) corresponds to requiring that $f(\rho)$ is a non-convex function of ρ . The values of ρ_1 and ρ_2 are obtained by minimizing $\gamma f(\rho'_1) + (1 - \gamma)f(\rho'_2)$ over all possible values of ρ'_1 and ρ'_2 , with γ determined by the mass conservation $\gamma\rho'_1 + (1 - \gamma)\rho'_2 = \rho_0$. The corresponding geometrical construction is called the common tangent construction (see Fig. 3.1).

We now try to translate the convexity condition (3.11) into a condition on the utility function $u(\rho)$. Phase separation occurs if there is a range of density for which

Fig. 3.1 Sketch of phase separation: the system of density ρ_0 splits into two phases of densities ρ_1 and ρ_2 to lower its free energy. The densities ρ_1 and ρ_2 are determined from the common tangent construction (dashed line) on the free energy curve $f(\rho)$ (full line). The free energy of the phase separated system is given by the value of the tangent at $\rho = \rho_0$



$f(\rho)$ is concave, namely $f''(\rho) < 0$. We thus compute the second derivative of f , yielding

$$f''(\rho) = \frac{T}{\rho(1-\rho)} - u'(\rho). \quad (3.12)$$

For a given utility function, the sign of $f''(\rho)$ can be checked explicitly. We note that in the limit $T \rightarrow 0$, $f''(\rho) = -u'(\rho)$, so that the homogeneous state is stable (i.e., $f''(\rho) > 0$) if $u(\rho)$ is a monotonously decreasing function of ρ .

The specific form of the utility function is an input of the model, and it can be postulated on a phenomenological basis, or rely on a theory of the interactions among agents. In order to analyze an explicit example of a non-linear utility function, we consider the peaked utility function defined as:

$$u(\rho) = \begin{cases} 2\rho & \text{if } \rho \leq \frac{1}{2} \\ 2(1-\rho) & \text{if } \rho > \frac{1}{2} \end{cases} \quad (3.13)$$

which is maximum for $\rho = \frac{1}{2}$ (see left panel of Fig. 3.2). The expression of $f(\rho)$ can be easily deduced from $u(\rho)$, and is illustrated on the right panel of Fig. 3.2 for different values of T . To study the stability of the homogeneous phase, we look at the sign of $f''(\rho)$. One has for $\rho < 1/2$

$$f''(\rho) = \frac{T}{\rho(1-\rho)} - 2, \quad (3.14)$$

and for $\rho > 1/2$:

$$f''(\rho) = \frac{T}{\rho(1-\rho)} + 2. \quad (3.15)$$

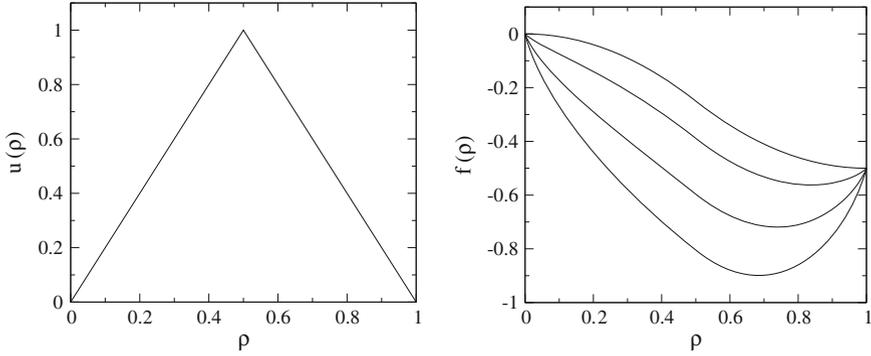


Fig. 3.2 *Left* utility function defined in Eq. (3.13). *Right* corresponding effective free energy $f(\rho)$, for different values of temperature, $T = 0, 0.2, 0.5$ and 0.8 (from *top* to *bottom*), illustrating that $f(\rho)$ becomes non-convex for $T < 0.5$, which leads to a phase separation

It is easy to check that $f''(\rho)$ is minimum for $\rho \rightarrow \frac{1}{2}^-$, the corresponding value being

$$\lim_{\rho \rightarrow \frac{1}{2}^-} f''(\rho) = 4T - 2. \quad (3.16)$$

Thus, for $T > 1/2$, the function $f(\rho)$ is convex on the whole interval $0 < \rho < 1$ as $f''(\rho) > 0$ on this interval, and the homogeneous phase is stable. On the contrary, for $T < 1/2$, there exists an interval of density ρ where $f(\rho)$ is concave ($f''(\rho) < 0$), so that in the stationary state, the system is split into two phases with different densities.

The surprising phenomenon here is that a phase separation occurs even in the case $\rho = 1/2$, although all agents have a significant preference for a half-filled neighborhood. This can be understood intuitively as follows. At a small, but non-zero temperature T , small fluctuations of density in the blocks are possible. Let us assume that we start from the homogeneous state of density $\rho = 1/2$, with some small fluctuations of this density around the mean value $1/2$. If a block has a density smaller than $1/2$, then this block becomes less attractive for the agents living in it. So some agents will start to move to the most attractive blocks which have exactly the density $1/2$. In doing so, the initial block becomes less and less attractive, thus making more and more agents leave it. This avalanche process, which is related to the selfish behavior of the agents, qualitatively explains the instability of the homogeneous state with density $1/2$.

Interestingly, this model can be slightly generalized to take into account a degree of ‘altruism’ of the agents, in the sense that agents may also partially take into account the cost imposed by their moves to the neighboring agents. This can be done by replacing Δu in Eq. (3.1) by a cost

$$C = \Delta u + \alpha(\Delta U - \Delta u) \quad (3.17)$$

where $0 \leq \alpha \leq 1$ is a weighting parameter, and ΔU is the total variation of utility of all agents in the city. The effective free energy $f(\rho)$ given in Eq. (3.10) is then changed into

$$f(\rho) = T\rho \ln \rho + T(1 - \rho) \ln(1 - \rho) - \alpha\rho u(\rho) - (1 - \alpha) \int_0^\rho u(\rho') d\rho'. \quad (3.18)$$

The case $\alpha = 1$ is actually very similar to what happens in physics, if one maps the agents' utility onto the opposite of an energy. The introduction of this parameter α has important consequences on the dynamics. It can be shown in particular that there exists a threshold value α_c such that for $\alpha > \alpha_c$, the phase separation disappears and the system remains homogeneous in steady state [3].

3.1.3 The 'True' Schelling Model: Two Types of Agents

To get closer to Schelling's original model, we briefly mention the case where two types of agents are included. We thus introduce agents of two colors such as 'red' and 'green'. The two types of agents are labelled with subindexes 'R' and 'G' in the following. To keep the model solvable, we assume that each type of agent is only concerned about the density of agents of the same type in its neighborhood. Hence the utility functions depend only on the density of the same type of agents, namely $u_R(\rho_R)$ and $u_G(\rho_G)$. We consider here the original 'selfish' dynamics corresponding to $\alpha = 0$. The effective free energy defined in Eq. (3.10) can be generalized to

$$\begin{aligned} f(\rho_R, \rho_G) &= T\rho_R \ln \rho_R + T\rho_G \ln \rho_G \\ &+ T(1 - \rho_R - \rho_G) \ln(1 - \rho_R - \rho_G) \\ &- \alpha [\rho_R u_R(\rho_R) + \rho_G u_G(\rho_G)] \\ &- (1 - \alpha) \left[\int_0^{\rho_R} u_R(\rho') d\rho' + \int_0^{\rho_G} u_G(\rho') d\rho' \right]. \end{aligned} \quad (3.19)$$

In order to determine the equilibrium configurations of the model, one needs to find the set $\{\rho_{qR}, \rho_{qG}\}$ minimizing the potential

$$F(\rho_{1R}, \dots, \rho_{QR}, \rho_{1G}, \dots, \rho_{QG}) = \sum_q f(\rho_{qR}, \rho_{qG}) \quad (3.20)$$

under the constraints that the total number of agents of each type is fixed, that is $\sum_q \rho_{qR} = Q\rho_{0R}$ and $\sum_q \rho_{qG} = Q\rho_{0G}$, where ρ_{0G} and ρ_{0R} are respectively the overall densities of 'green' and 'red' agents.

Due to the constraint that the total density of agents (disregarding their type) has to remain less than one, this model including two types of agents does not reduce to two uncoupled models composed of a single type of agents. It is however possible

to compute the stationary states. We consider again the small ‘temperature’ limit $T \rightarrow 0$, and assume for simplicity that the overall densities of ‘red’ and ‘green’ agents are equal, $\rho_{0R} = \rho_{0G} = \rho_0/2$. One then finds a segregated state in which each block contains a single type of agents, at density ρ_0 [3].

3.2 Driven Particles on a Lattice: Zero-Range Process

Let us now turn to a different type of situation, involving alternative techniques. While the above variant of the Schelling model could be dealt with by using a mapping to an equilibrium system, in many cases such equilibrium methods are not sufficient to solve the model, due for instance to the presence of fluxes in the system. One must then resort to other kinds of approaches. Among possible approaches, one can consider simple enough stochastic models for which an exact solution of the master equation can be found in the steady state, although detailed balance is not satisfied. A prominent example of such type of models is the so-called Zero-Range Process (ZRP) [5], that we describe below. Another well-known example of exactly solvable non-equilibrium model is the Asymmetric Simple Exclusion Process (ASEP), for which the derivation of the solution is however much more technical [6].

3.2.1 Definition and Exact Steady-State Solution

In the ZRP, N particles are randomly placed on the L sites of a one-dimensional lattice with periodic boundary conditions,² and can jump from site i to the neighboring site $i + 1$ (with the convention $L + 1 \equiv 1$). Motion is thus biased, which generates a current of particles along the ring. The interaction between particles is taken into account through the fact that the probability per unit time to jump from site i to site $i + 1$ depends on the current number n_i of particles on site i ; this probability is denoted as $u(n_i)$.

A configuration of the ZRP is given by the set $C = (n_1, \dots, n_L)$ of the occupation numbers of all sites. The transition rate $W(C'|C)$ can be written formally as

$$W(n'_1, \dots, n'_L | n_1, \dots, n_L) = \sum_{i=1}^L u(n_i) \delta_{n'_i, n_i-1} \delta_{n'_{i+1}, n_{i+1}+1} \prod_{j \neq i, i+1} \delta_{n'_j, n_j} \quad (3.21)$$

where $\delta_{n', n}$ is the Kronecker symbol, equal to 1 if $n' = n$, and to 0 otherwise. Using this form of the transition rate, one can write the corresponding master equation (see

²We consider here for simplicity the ring geometry, but the ZRP can actually be defined on an arbitrary graph [7].

Sect. 2.1), which we do not display here to lighten the presentation. It can be shown [5] that the steady-state distribution takes a factorized form

$$P(n_1, \dots, n_L) = \frac{1}{Z} \left(\prod_{i=1}^L f(n_i) \right) \delta_{\sum_j n_j, N} \quad (3.22)$$

where the Kronecker delta symbol accounts for the conservation of the total number of particles. Inserting this form into the master equation, one obtains the expression of $f(n)$:

$$f(n) = \begin{cases} \prod_{k=1}^n \frac{1}{u(k)} & \text{if } n \geq 1, \\ 1 & \text{if } n = 0. \end{cases} \quad (3.23)$$

Note that the model can also be defined in such a way as to obtain any desired function $f(n)$ in the steady-state distribution: one simply needs to choose $u(n) = f(n-1)/f(n)$, for $n \geq 1$.

3.2.2 Maximal Density and Condensation Phenomenon

One of the interesting properties of the ZRP is the presence of a condensation transition, where a finite fraction of the total number of particles gather on a single site. Such a phenomenon appears in the case of a function $f(n)$ decaying as a power-law, $f(n) \sim 1/n^\alpha$, or equivalently $u(n) = 1 + \alpha/n + o(1/n)$. The single-site distribution can be obtained by considering the rest of the system as a reservoir of particles, a situation similar to the canonical ensemble at equilibrium. Assuming the system to be homogeneous, the single-site distribution is then given by

$$p(n) = c f(n) e^{-\mu n} \quad (3.24)$$

where μ is the effective chemical potential of the reservoir. The normalization constant c is determined by

$$\frac{1}{c} = \sum_{n=0}^{\infty} f(n) e^{-\mu n}. \quad (3.25)$$

The convergence of this sum requires that $\mu > 0$ (or $\mu \geq 0$ if $\alpha > 1$). The average density

$$\rho = \langle n \rangle = c \sum_{n=1}^{\infty} n f(n) e^{-\mu n} \quad (3.26)$$

is a decreasing function of μ , which thus reaches its maximum value ρ_c for $\mu \rightarrow 0$:

$$\rho_c = c \sum_{n=1}^{\infty} n f(n) \sim \sum_{n=1}^{\infty} \frac{1}{n^{\alpha-1}}. \quad (3.27)$$

Hence ρ_c is infinite if $\alpha \leq 2$, and finite if $\alpha > 2$. As a result, if $\alpha > 2$, a homogeneous density of particles cannot exceed a finite density ρ_c . If, on the contrary, one imposes a density $\rho_0 > \rho_c$, by including in the system a number of particles $N > L\rho_c$, the dynamics will necessarily evolve toward a non-homogeneous state. It can be shown [5] that the resulting state is composed of a ‘fluid phase’, homogeneous at density ρ_c , and a ‘condensate’, that is a single site containing a macroscopic number of particles $L(\rho_0 - \rho_c)$.

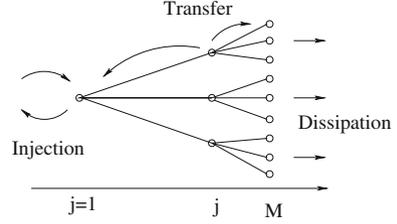
Applications of this model range from vibrated granular matter (each site corresponding to a vibrated urn containing grains), to road traffic (sites being sections of roads), or network dynamics (that is, the dynamics of attachment and detachment of links on the network) [5]. Note that the ZRP is a very simplified model, so that mapping it to more realistic situations often implies approximations.

3.2.3 Dissipative Zero-Range Process

To go beyond the simple Zero Range Process discussed above, it is interesting to consider situations where on the one hand the geometry is more complex than the simple one-dimensional ring, and on the other hand particles are exchanged with reservoirs. The case of a single reservoir can be treated in a simple way by replacing the Kronecker δ in Eq. (3.22) by an exponential factor $\exp(-\mu \sum_j n_j)$, where μ is the chemical potential of the reservoir. One finds in this case that the system remains homogeneous for all values of the chemical potential: the condensation transition occurs only in the case when the total number of particles is fixed. A more interesting situation corresponds to the case of two reservoirs. Then, depending on the properties of the reservoirs, a steady flux of particles may be observed from one reservoir to the other. Alternatively, one may also interpret the particles as fixed amounts of energy that can be exchanged between different nodes of the lattice—the chemical potential is then replaced by an inverse temperature. This is the interpretation we retain below, where we combine a non-trivial geometry (a tree geometry) with the presence of two reservoirs.

The model is defined as follows [8]. The network on which the dynamics takes place is a tree composed of M successive levels. At each level $j < M$, each site has a number $m > 1$ of forward branches linked to a node at level $j + 1$ —see Fig. 3.3. The number of nodes at a level j is equal to m^{j-1} . Nodes are thus labeled by their level index j , and by a further index $i = 1, \dots, m^{j-1}$ distinguishing the different nodes present at the same level. It is convenient to think of each level of the tree as a different length scale. Large length scales correspond to the top of the tree, and

Fig. 3.3 Sketch of the dissipative model, illustrating the tree geometry in the case of $M = 3$ levels and $m = 3$ forward branches per node



include a small number of degrees of freedom. On the contrary, the bottom of the tree describes small length scales, and is associated to a large number of degrees of freedom. To associate a length scale to each level j of the tree, it is convenient to also introduce a quantity $k_j = m^{j-1}$, interpreted as a ‘pseudo-wave number’ in a physics terminology. The corresponding length scale is then given by $\ell_j = 1/k_j$.

By connecting a high-temperature reservoir at large scales, and a low-temperature one at small scales, the resulting model may be thought of as a dissipative system, where energy is injected at large scales, and dissipated at small scales. Physical examples of this broad class of systems include for instance hydrodynamic turbulence [9], wave turbulence in fluids or in plasma [10], and vibrating plates [11, 12]. Of course, the dissipative ZRP that we consider here is only a toy model, that cannot account for all the complexity of these realistic systems.

We assume that each node can carry an energy taking only discrete values proportional to a finite amount ε_0 , so that the energy $\varepsilon_{j,i}$ at node (j, i) is equal to $n_{j,i}\varepsilon_0$, with $n_{j,i}$ an integer. The configuration of the system is thus described by the list of all the $n_{j,i}$ ’s. Energy transfer within the tree proceeds by moving an amount ε_0 of energy along any branch linking level j and $j + 1$ with a rate (probability per unit time) $\nu_j = \nu k_j^\alpha$, where ν is a constant parameter. Energy injection is modeled by connecting an energy reservoir at temperature $T_{\text{ext}} = \beta_{\text{ext}}^{-1}$ to the level $j = 1$. The frequency of exchanges is chosen to be equal to ν for simplicity (but the model can be equally solved for an arbitrary value of this coupling to the reservoir). To model dissipation, we assume that energy is randomly withdrawn from any node at level M with a rate Δ_M .

By solving the associated master equation, it can be shown that the stationary probability distribution $P_{\text{st}}(\{n_{j,i}\})$ takes the form

$$P_{\text{st}}(\{n_{j,i}\}) = \frac{1}{Z} \prod_{j=1}^M \prod_{i=1}^{m^{j-1}} e^{-\beta_j n_{j,i} \varepsilon_0} \quad (3.28)$$

where the parameters β_j are effective inverse temperatures associated to level j of the tree, and Z is a normalization factor. For the distribution $P_{\text{st}}(\{n_{j,i}\})$ to solve the master equation of the model, the inverse temperatures β_j have to satisfy the following set of equations, expressed in terms of the parameters $z_j = \exp(-\beta_j \varepsilon_0)$:

$$\nu_{j-1}(z_{j-1} - z_j) - m\nu_j(z_j - z_{j+1}) = 0, \quad j = 2, \dots, M-1 \quad (3.29)$$

with the boundary conditions

$$\nu(e^{-\beta_{\text{ext}}\varepsilon_0} - z_1) - m\nu_1(z_1 - z_2) = 0, \quad (3.30)$$

$$\nu_{M-1}(z_{M-1} - z_M) = \Delta_M z_M. \quad (3.31)$$

These equations can also be interpreted as the local balance of the diffusive currents $\nu_j(z_j - z_{j+1})$ and the dissipative current $\Delta_M z_M$. Note that if the dissipation Δ_M is set to zero, no more current flows in the system, leading to an equilibrium solution with $\beta_1 = \dots = \beta_M = \beta_{\text{ext}}$. Solving Eqs. (3.29)–(3.31), two different temperature profiles may be obtained in the large M limit depending on the value of α . For $\alpha < -1$, the temperature profile slowly converges to the constant value $\beta_j = \beta_{\text{ext}}$ imposed by the reservoir. In contrast, for $\alpha > -1$, the temperature profile converges in the large M limit to a genuine nonequilibrium profile given by

$$\beta_j^{\text{neq}} = \frac{1}{\varepsilon_0}(1 + \alpha) \ln k_j + \beta_{\text{ext}} + \frac{1}{\varepsilon_0} \ln c, \quad (3.32)$$

with $c = 1 + m - m^{-\alpha}$. Note that the nonequilibrium profile β_j^{neq} is determined only by parameters characterizing energy injection and transfer, and not by parameters related to the dissipation mechanism. Interestingly, the temperature profile is continuous as a function of α , in the sense that $\beta_j \rightarrow \beta_{\text{ext}}$ when $\alpha \rightarrow -1^+$.

To better understand the origin of this change of behavior around $\alpha = -1$, one can compute the mean energy flux Φ crossing the system,

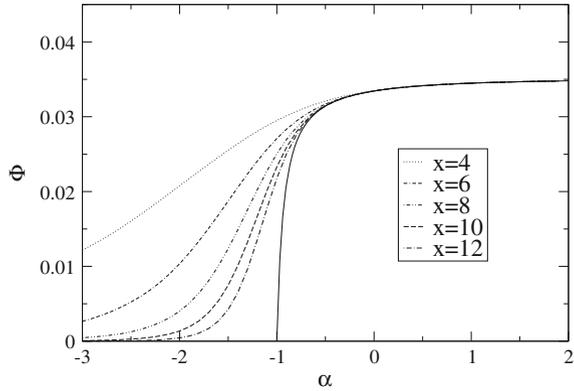
$$\Phi = \nu(e^{-\beta_{\text{ext}}\varepsilon_0} - e^{-\beta_1\varepsilon_0}). \quad (3.33)$$

In the limit $M \rightarrow \infty$, this flux is found to be

$$\Phi = \frac{\nu}{c}(m - m^{-\alpha})e^{-\beta_{\text{ext}}\varepsilon_0} \quad (3.34)$$

for $\alpha > -1$ and $\Phi = 0$ for $\alpha < -1$. The energy flux Φ is plotted as a function of α in Fig. 3.4, showing a continuous transition at the value $\alpha = -1$. Hence the emergence of a non-uniform temperature profile across scales is associated to the existence of a finite injected and dissipated energy flux across the system. To sum up, the system reaches a quasi-equilibrium state in the large size limit for $\alpha < -1$, while it reaches a true non-equilibrium state, with a finite energy flux, for $\alpha > -1$. A peculiar feature of this non-equilibrium state is that it explicitly depends on the internal dynamics characterized by α , as can be seen on Eq. (3.32). In contrast, the quasi-equilibrium state obtained for $\alpha < -1$ does not depend on α (at least asymptotically, for $M \rightarrow \infty$); it is only characterized by the temperature β_{ext} of the reservoir.

Fig. 3.4 Dissipated flux Φ as a function of α , for different values $D = 10^{-x}$ of the dissipation coefficient. Parameters: $m = 2$, $\nu = 0.1$, $\beta_{\text{ext}} = 1$, $\varepsilon_0 = 1$



We have thus seen that the presence of a non-vanishing flux is a key ingredient to generate a nonequilibrium steady state. At a qualitative level, the need for a flux to sustain a nonequilibrium steady-state may be considered as a relatively general feature of complex systems, which is not specific to the physical model considered here, and may be found in very different contexts, like for instance in economics. As an elementary example, a store cannot maintain its activity if the flux of sold products is too low. This can be understood within the framework of a very simplified model of store, in which only one type of product is sold. Let us call ϕ the flux of products, that is, the number of products sold per unit time. The total profit made by the store is simply $P = \phi \Delta p - C_0$, where Δp is the difference between buying and selling prices of products, and C_0 is the fixed cost per unit time for running the store (rent for the building, employee wage, etc.). Of course, the activity of the store can be maintained only if the profit P is positive. As a result, ϕ has to be larger than $\phi_c = C_0/\Delta p$. Hence a store is a non-equilibrium system which needs a flux (of sold product) larger than a threshold value to maintain a steady state.

3.3 Collective Motion of Active Particles

Active particles are particles able to sustain a continuous motion thanks to some external energy input. This concept is used by physicists to describe for instance the motion of animals, bacteria, or more recently different types of self-driven colloids [13]. A very schematic model of active particle is a point-like particle with a velocity vector of constant modulus, but arbitrary direction. In the simplest cases, the direction of motion of the particles just diffuses randomly, and one speaks about active Brownian particles. Different types of interactions may be included between active particles, like repulsion forces for instance. In this case, the interplay between self-propulsion and repulsion leads to a phase separation, with the formation of dense clusters, as if the particles had effective attractive interactions [14]. Besides, other types of inter-

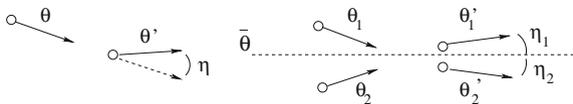


Fig. 3.5 Schematic representation of the dynamics of the model. *Left* self-diffusion of the velocity angle; *Right* binary collision. See text for notations

actions which are specific to self-propelled particles can be included. This is the case in particular of velocity-alignment interactions. A paradigmatic model for this type of interactions between active particles is the so-called Vicsek model, that has been extensively studied through numerical simulations [15–17]. A transition from disordered motion when the density of active particles is low, to ordered collective motion when the density is high, has been reported. This transition exhibits some properties similar to that of phase transitions observed in physical systems. It is also possible to develop analytical approaches, either by postulating phenomenological equations of motion at the macroscopic scale (hydrodynamic equations) [18], or by using a Boltzmann approach to derive such hydrodynamic equations [19]. We present here a brief summary of the results obtained from the latter approach.

We consider self-propelled point-like particles moving on a continuous two-dimensional space, with a velocity vector \mathbf{v} of fixed magnitude v_0 (to be chosen as the speed unit) in a reference frame. The velocity of the particles is simply defined by the angle θ between \mathbf{v} and a given reference direction. Particles move in straight line, following their velocity vector, until they experience either a self-diffusion event (a random scattering), or a binary collision that tends to align the velocities of the two particles—see Fig. 3.5. Self-diffusion events are defined as follows: the velocity angle θ of any particle is changed with a probability λ per unit time to a value $\theta' = \theta + \eta$, where η is a Gaussian noise with distribution $p_0(\eta)$ and variance σ_0^2 . Binary collisions, that are the only interactions between particles, occur when the distance between two particles becomes less than d_0 (in the following, we set $d_0 = \frac{1}{2}$). The velocity angles θ_1 and θ_2 of the two particles are then changed into $\theta'_1 = \bar{\theta} + \eta_1$ and $\theta'_2 = \bar{\theta} + \eta_2$, as shown on Fig. 3.5. In the last expression, $\bar{\theta} = \arg(e^{i\theta_1} + e^{i\theta_2})$ is the average angle, and η_1 and η_2 are independent Gaussian noises with the same distribution $p(\eta)$ and variance σ^2 . Note that these binary collisions are different from the collisions in usual gases, as in this latter case, collisions are ruled by energy and momentum conservation laws. In the following, we take for simplicity identical distributions $p_0(\eta)$ and $p(\eta)$; a single parameter σ thus characterizes the amplitude of the noise.

3.3.1 Derivation of Continuous Equations

A useful mathematical tool to describe statistically the dynamics of the system is the one-particle phase-space distribution $f(\mathbf{r}, \theta, t)$, namely the probability to find

a particle at position \mathbf{r} and with a velocity angle θ , at time t . The evolution of this one-particle phase-space distribution is ruled by the Boltzmann equation, which reads

$$\frac{\partial f}{\partial t}(\mathbf{r}, \theta, t) + \mathbf{e}(\theta) \cdot \nabla f(\mathbf{r}, \theta, t) = I_{\text{dif}}[f] + I_{\text{col}}[f]. \quad (3.35)$$

The functionals $I_{\text{dif}}[f]$ and $I_{\text{col}}[f]$ respectively account for the self-diffusion and collision phenomena. The vector $\mathbf{e}(\theta)$ is the unit vector in the direction θ . The diffusion functional $I_{\text{dif}}[f]$ is given by

$$I_{\text{dif}}[f] = -\lambda f(\mathbf{r}, \theta, t) + \lambda \int_{-\infty}^{\infty} d\eta p(\eta) f(\mathbf{r}, \theta - \eta, t). \quad (3.36)$$

The evaluation of the collision term $I_{\text{col}}[f]$ is more subtle. We know that two particles collide if their distance becomes less than the interaction range d_0 . In the frame of particle 1, particle 2 has a velocity $\mathbf{v}'_2 = \mathbf{e}(\theta_2) - \mathbf{e}(\theta_1)$. Hence, particles that collide with particle 1 between t and $t + dt$ are those that lie, at time t , in a rectangle of length $|\mathbf{v}'_2| dt$ and of width $2d_0$, yielding for the collision functional (the collision area does not change going back to the lab frame)

$$\begin{aligned} I_{\text{col}}[f] &= -f(\mathbf{r}, \theta, t) \int_{-\pi}^{\pi} d\theta' |\mathbf{e}(\theta') - \mathbf{e}(\theta)| f(\mathbf{r}, \theta', t) \\ &\quad + \int_{-\pi}^{\pi} d\theta_1 \int_{-\pi}^{\pi} d\theta_2 \int_{-\infty}^{\infty} d\eta p(\eta) |\mathbf{e}(\theta_2) - \mathbf{e}(\theta_1)| \\ &\quad \times f(\mathbf{r}, \theta_1, t) f(\mathbf{r}, \theta_2, t) \delta_{2\pi}(\bar{\theta} + \eta - \theta), \end{aligned} \quad (3.37)$$

with $\bar{\theta} = \arg(e^{i\theta_1} + e^{i\theta_2})$, and $\delta_{2\pi}$ a generalized Dirac distribution taking into account the periodicity of angles. One can check that the uniform angular distribution $f(\mathbf{r}, \theta, t) = \rho/2\pi$ is a solution of Eq. (3.35) for an arbitrary constant density ρ , and for any value of the noise amplitude σ .

In order to deal with convenient physical quantities, we introduce the hydrodynamic density and velocity fields $\rho(\mathbf{r}, t)$ and $\mathbf{u}(\mathbf{r}, t)$:

$$\rho(\mathbf{r}, t) = \int_{-\pi}^{\pi} d\theta f(\mathbf{r}, \theta, t), \quad (3.38)$$

$$\mathbf{u}(\mathbf{r}, t) = \frac{1}{\rho(\mathbf{r}, t)} \int_{-\pi}^{\pi} d\theta f(\mathbf{r}, \theta, t) \mathbf{e}(\theta). \quad (3.39)$$

Integrating the Boltzmann equation (3.35) over θ , one directly obtains the continuity equation for $\rho(\mathbf{r}, t)$:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0. \quad (3.40)$$

The operator ∇ is the vectorial differential operator³ of components $(\partial/\partial x, \partial/\partial y)$. The derivation of a hydrodynamic equation for the velocity field is less straightforward, and involves an approximation scheme. The reader is referred to Refs. [19, 20] for more details on the derivation. The principle of the derivation is to expand the distribution $f(\mathbf{r}, \theta, t)$ into angular Fourier modes according to⁴

$$f(\mathbf{r}, \theta, t) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} f_k(\mathbf{r}, t) e^{-ik\theta}, \quad (3.41)$$

where the Fourier coefficients f_k are defined as

$$f_k(\mathbf{r}, t) = \int_{-\pi}^{\pi} d\theta f(\mathbf{r}, \theta, t) e^{ik\theta}. \quad (3.42)$$

Note that $f_0(\mathbf{r}, t)$ is nothing but the local density $\rho(\mathbf{r}, t)$. Note also that for all k , $f_{-k} = f_k^*$ (the star denotes the complex conjugate), since $f(\mathbf{r}, \theta, t)$ is real.

The Boltzmann equation can in turn be expanded into Fourier modes, leading to

$$\frac{\partial f_k}{\partial t} + \frac{v_0}{2} (\partial f_{k-1} + \partial^* f_{k+1}) = -(1 - P_k) f_k + \sum_{q=-\infty}^{\infty} J_{k,q} f_q f_{k-q} \quad (3.43)$$

where we have used the shorthand notation ∂ and ∂^* for the complex differential operators

$$\partial \equiv \frac{\partial}{\partial x} + i \frac{\partial}{\partial y}, \quad \partial^* \equiv \frac{\partial}{\partial x} - i \frac{\partial}{\partial y}. \quad (3.44)$$

The coefficient $J_{k,q}$ is given by $J_{k,q} = P_k(\sigma) I_{k,q} - I_{0,q}$, where $I_{k,q}$ is defined by the integral

$$I_{k,q} = \frac{1}{\pi} \int_{-\pi}^{\pi} dx \left| \sin \frac{x}{2} \right| e^{-iqx + ikx/2} \quad (3.45)$$

with $P_k(\sigma) = \int_{-\infty}^{\infty} d\eta P_\sigma(\eta) e^{ik\eta}$ the Fourier transform of the noise distribution (restricted here to integer values of k). One has $0 \leq P_k(\sigma) \leq 1$ and $P_k(0) = 1, \forall k$. For a Gaussian noise distribution, the Fourier transform has the simple form

$$P_k(\sigma) = e^{-\sigma^2 k^2 / 2}. \quad (3.46)$$

³More explicitly, Eq. (3.40) reads

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u_x) + \frac{\partial}{\partial y}(\rho u_y) = 0,$$

where (u_x, u_y) are the components of the vector \mathbf{u} .

⁴Here, i is a complex number such that $i^2 = -1$.

To proceed further, it is necessary to identify the linear instability of the disordered state, corresponding to an isotropic distribution for which $f_k = 0$ for all $k \neq 0$. This linear instability occurs for a critical value ρ_c which depends on the intensity of the noise. For an average density ρ_0 slightly above ρ_c , one assumes that the distribution $f(\mathbf{r}, \theta, t)$ is still close to an isotropic distribution, and makes a scaling ansatz for the different fields at stake, as a function of a small parameter ϵ characterizing the distance to the threshold density,

$$f_1 \sim \epsilon, \quad f_2 \sim \epsilon^2, \quad \rho - \rho_0 \sim \epsilon. \quad (3.47)$$

One also needs to make similar assumptions about the space and time derivatives. A coherent scaling ansatz turns out to be

$$\partial \sim \epsilon, \quad \frac{\partial}{\partial t} \sim \epsilon, \quad (3.48)$$

consistently with the propagative nature of the dynamics. The Boltzmann equation, expressed in Fourier modes, is then truncated to order ϵ^3 , neglecting terms of order ϵ^4 or higher. This results in two coupled equations for f_1 and f_2 . The Fourier coefficient f_1 is a complex number having real and imaginary parts equal to components of the pseudo-momentum field $\mathbf{w}(\mathbf{r}, t) = \rho(\mathbf{r}, t)\mathbf{u}(\mathbf{r}, t)$. Hence this is the relevant field we need to keep in the description to characterize the emergence of polar order. The second Fourier coefficient f_2 actually has a much faster relaxation dynamics than f_1 . Using this separation of time scales, one can express f_2 as a function of f_1 , and obtain in this way a closed equation for the evolution of f_1 (which is however still coupled to the density ρ , the other relevant field).

Mapping complex numbers onto vectors, we end up with the following hydrodynamic equations:

$$\begin{aligned} \frac{\partial \mathbf{w}}{\partial t} + \gamma(\mathbf{w} \cdot \nabla)\mathbf{w} &= -\frac{1}{2}\nabla(\rho - \kappa\mathbf{w}^2) \\ &+ (\mu - \xi\mathbf{w}^2)\mathbf{w} + \nu\Delta\mathbf{w} - \kappa(\nabla \cdot \mathbf{w})\mathbf{w} \end{aligned} \quad (3.49)$$

with Δ the Laplacian operator,

$$\Delta \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \quad (3.50)$$

It is interesting to give a physical interpretation of the different terms appearing in this hydrodynamic equation. The first term in the r.h.s. of Eq. (3.49) can be interpreted as a pressure gradient, considering $p = \frac{1}{2}(\rho - \kappa\mathbf{w}^2)$ as an effective pressure. The second term accounts for the local relaxation of \mathbf{w} , while the third term is analogous to the standard viscous term appearing in the Navier-Stokes equation describing usual fluids. Finally, the last term corresponds to a feedback on the flow from compressibility effects.

The different coefficients appearing in Eq. (3.49) can be computed explicitly as a function of the microscopic parameters of the model. They are given by [19]

$$\nu = \frac{1}{4} \left[\lambda \left(1 - e^{-2\sigma_0^2} \right) + \frac{4}{\pi} \rho \left(\frac{14}{15} + \frac{2}{3} e^{-2\sigma^2} \right) \right]^{-1}, \quad (3.51)$$

$$\gamma = \frac{8\nu}{\pi} \left(\frac{16}{15} + 2e^{-2\sigma^2} - e^{-\sigma^2/2} \right), \quad (3.52)$$

$$\kappa = \frac{8\nu}{\pi} \left(\frac{4}{15} + 2e^{-2\sigma^2} + e^{-\sigma^2/2} \right), \quad (3.53)$$

$$\mu = \frac{4}{\pi} \rho \left(e^{-\sigma^2/2} - \frac{2}{3} \right) - \lambda \left(1 - e^{-\sigma_0^2/2} \right), \quad (3.54)$$

$$\xi = \frac{64\nu}{\pi^2} \left(e^{-\sigma^2/2} - \frac{2}{5} \right) \left(\frac{1}{3} + e^{-2\sigma^2} \right). \quad (3.55)$$

Note that ν , γ and κ are always positive; μ can change sign, and $\xi > 0$ whenever $\mu > 0$.

3.3.2 Phase Diagram and Instabilities

Turning to the study of the spontaneous onset of collective motion in the present model, we look for possible instabilities of the spatially homogeneous flow, that is the appearance of a uniform, nonzero, velocity field \mathbf{u} (or pseudo-momentum field \mathbf{w}). Considering a time-dependent, but spatially homogeneous flow, we get

$$\frac{\partial \mathbf{w}}{\partial t} = (\mu - \xi \mathbf{w}^2) \mathbf{w}. \quad (3.56)$$

Obviously, $\mathbf{w} = 0$ is a solution for arbitrary values of the coefficients. However, this solution becomes unstable for $\mu > 0$, when a nonzero solution $\mathbf{w}_0 = \sqrt{\mu/\xi} \mathbf{e}$ appears (\mathbf{e} is a unit vector pointing in a arbitrary direction). From the expression (3.54) of μ , it turns out that $\mu = 0$ corresponds to a threshold value ρ_t given by

$$\rho_t = \frac{\pi \lambda (1 - e^{-\sigma_0^2/2})}{4(e^{-\sigma^2/2} - \frac{2}{3})}. \quad (3.57)$$

The transition line defined by ρ_t in the plane (ρ, σ) is plotted on Fig. 3.6. The instability is seen to occur at any density, provided the noise is low enough. The transition line saturates at a value $\sigma_t = (2 \ln \frac{3}{2})^{1/2} \approx 0.90$.

Further instabilities leading to more complicated patterns, like travelling solitary waves are also observed, both at the level of the hydrodynamic equations [19] and in numerical simulations of the Vicsek model [16]. These instabilities occur in the parameter region denoted as **B** in Fig. 3.6.

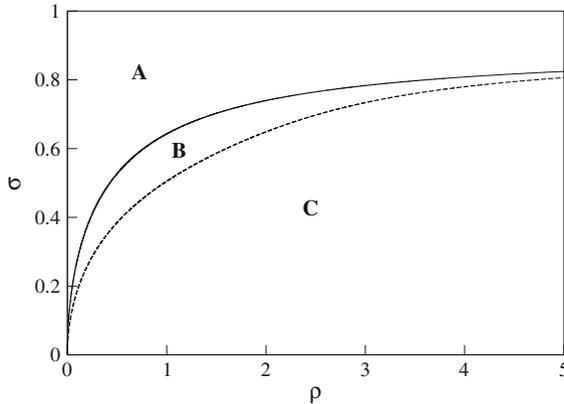


Fig. 3.6 Phase diagram of the hydrodynamic equation (3.49) in the noise-density plane ($\lambda = 1$, $d_0 = 0.5$, $v_0 = 1$). A transition line (full line) separates the domains with zero hydrodynamic velocity (region A), from the domain where collective motion occurs (regions B and C). In region C, homogeneous motion is stable (except if one goes to high density and low noise, in which case a further instability—not shown here—appears). In region B, homogeneous motion is unstable, and one observes solitary waves of high density moving over a disordered background

3.3.3 Varying the Symmetries of Particles

A similar approach can be used in models having different symmetries from the ones considered here. For instance, considering an experiment with vibrated rice grains, one observes that these vibrated elongated grains have a tendency to move back and forth along their main axis preferentially. Similarly to the self-propelled particles considered in the previous section, such particles can also be attributed a direction characterized by an angle θ (we are still considering a two-dimensional problem), but their internal symmetry renders the directions θ and $\theta + \pi$ equivalent. Interactions between grains also tend to align closeby grains, but they need to take into account the equivalence of the directions θ and $\theta + \pi$, called nematic symmetry.

To model such an experimental situation, we consider a model of point-like particles quite similar to the one introduced previously, but taking into account the nematic symmetry. At each elementary time step, particles move randomly either in the θ or in the $\theta + \pi$ direction with equal probability. Upon a collision, the angles θ_1 and θ_2 of the two particles are changed into $\theta'_1 = \bar{\theta} + \eta_1$ and $\theta'_2 = \bar{\theta} + \eta_2$, where now $\bar{\theta} = \arg(e^{i\tilde{\theta}_1} + e^{i\tilde{\theta}_2})$ is the average angle obtained by choosing $\tilde{\theta}_1 = \theta_1[\pi]$ and $\tilde{\theta}_2 = \theta_2[\pi]$ such that $|\tilde{\theta}_2 - \tilde{\theta}_1| < \frac{\pi}{2}$. As before, η_1 and η_2 are independent Gaussian noises with the same distribution $p(\eta)$ and variance σ^2 .

This model can be studied following the same lines [20, 21] as for the polar case by writing a Boltzmann equation, which takes a form similar to Eq. (3.35), but includes diffusion terms (of second order in space derivative) instead of the drift terms (of first order in space derivative), since particles are diffusing with no net motion. Integrating

the Boltzmann equation over the angles yields the following evolution equation for the density ρ (making an appropriate choice of units):

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \Delta \rho + \frac{1}{2} \text{Re}(\partial^{*2} f_2), \quad (3.58)$$

where f_2 is the complex field describing nematic order. To derive an evolution equation for this order parameter f_2 , a procedure similar to the one used in Sect. 3.3.1 is used. Namely, one expands the Boltzmann equation in Fourier space, identifies the linear instability threshold and determines a consistent scaling ansatz close to threshold. Note that quite importantly, all odd Fourier coefficients f_1, f_3 , etc. are equal to zero due to the nematic symmetry. The relevant order parameter is thus the second Fourier coefficient f_2 , and keeping the next Fourier coefficient f_4 in the truncation procedure turns out to be important to obtain the nonlinear terms saturating the instability. The relevant scaling ansatz reads in this case

$$f_2 \sim \epsilon, \quad f_4 \sim \epsilon^2, \quad \rho - \rho_0 \sim \epsilon, \quad \partial \sim \epsilon, \quad \frac{\partial}{\partial t} \sim \epsilon^2, \quad (3.59)$$

now corresponding to a diffusive space-time scaling, in line with the diffusive nature of the dynamics. After truncation and closure, one eventually obtains the following equation for the nematic field f_2

$$\frac{\partial f_2}{\partial t} = (\mu - \xi |f_2|^2) f_2 + \frac{1}{2} \Delta f_2 + \frac{1}{4} \partial^2 \rho, \quad (3.60)$$

where the coefficients μ and ξ can be expressed as a function of the density and of microscopic parameters of the models [21]. Here again, μ is negative below a threshold density $\rho_t(\sigma)$, and positive above. It is interesting to note that this equation is a generalization of the standard Ginzburg-Landau equation [22], with a coupling to the density field in the term linear in f_2 . The phase diagram of this equation is quite similar to that of the polar case, replacing polar order by nematic order. In particular, the linear instability line in the noise-density plane has a shape similar to that shown in Fig. 3.6 for the polar case. One of the main differences is that the high density travelling bands appearing in the polar case are replaced in the nematic case by high density bands that do not move, and that are now oriented along the axis of order (while polar order is perpendicular to the band in the polar case). Quite importantly, these bands are themselves unstable for large enough systems, leading to a regime of spatiotemporal chaos in which long-range nematic order cannot build up; order is only restored going to higher density or lower noise.

Note that other cases with ‘mixed’ symmetries can also be considered [20], for instance self-propelled particles interacting nematically, a case sometimes called ‘self-propelled rods’ with experimental realizations, e.g., in biological systems like colonies of bacteria [23].

References

1. S.F. Edwards, R.B.S. Oakeshott, Theory of powders. *Phys. A* **157**, 1080 (1989)
2. T.C. Schelling, Dynamic models of segregation. *J. Math. Sociol.* **1**, 143 (1971)
3. S. Grauwlin, E. Bertin, R. Lemoy, P. Jensen, Competition between collective and individual dynamics. *Proc. Natl. Acad. Sci.* **106**, 20622 (2009)
4. H. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, New York, 1985)
5. M.R. Evans, T. Hanney, Nonequilibrium statistical mechanics of the Zero-Range Process and related models. *J. Phys. A* **38**, R195 (2005)
6. B. Derrida, M.R. Evans, V. Hakim, V. Pasquier, Exact solution of a 1D asymmetric exclusion model using a matrix formulation. *J. Phys. A* **26**, 1493 (1993)
7. M.R. Evans, S.N. Majumdar, R.K.P. Zia, Factorized steady states in mass transport models on an arbitrary graph. *J. Phys. A* **39**, 4859 (2006)
8. E. Bertin, O. Dauchot, Far-from-equilibrium state in a weakly dissipative model. *Phys. Rev. Lett.* **102**, 160601 (2009)
9. U. Frisch, *Turbulence* (Cambridge University Press, Cambridge, 1995)
10. V.E. Zakharov, V.S. L'vov, G. Falkovich, *Kolmogorov Spectra of Turbulence I: Wave Turbulence* (Springer, Berlin, 1992)
11. G. Düring, C. Josserand, S. Rica, Weak turbulence for a vibrating plate: can one hear a Kolmogorov spectrum? *Phys. Rev. Lett.* **97**, 025503 (2006)
12. A. Boudaoud, O. Cadot, B. Odille, C. Touzé, Observation of wave turbulence in vibrating plates. *Phys. Rev. Lett.* **100**, 234504 (2008)
13. M.C. Marchetti, J.-F. Joanny, S. Ramaswamy, T.B. Liverpool, J. Prost, M. Rao, R.A. Simha, Soft active matter. *Rev. Mod. Phys.* **85**, 1143 (2013)
14. M.E. Cates, J. Tailleur, Motility-induced phase separation. *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015)
15. T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, O. Shochet, Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**, 1226 (1995)
16. G. Grégoire, H. Chaté, Onset of collective and cohesive motion. *Phys. Rev. Lett.* **92**, 025702 (2004)
17. J. Toner, Y. Tu, S. Ramaswamy, Hydrodynamics and phases of flocks. *Ann. Phys.* **318**, 170 (2005)
18. J. Toner, Y. Tu, Long-range order in a two-dimensional dynamical XY model: how birds fly together. *Phys. Rev. Lett.* **75**, 4326 (1995)
19. E. Bertin, M. Droz, G. Grégoire, Hydrodynamic equations for self-propelled particles: microscopic derivation and stability analysis. *J. Phys. A* **42**, 445001 (2009)
20. A. Peshkov, E. Bertin, F. Ginelli, H. Chaté, Boltzmann-Ginzburg-Landau approach for continuous descriptions of generic Vicsek-like models. *Eur. Phys. J. Spec. Topics* **223**, 1315 (2014)
21. E. Bertin, H. Chaté, F. Ginelli, S. Mishra, A. Peshkov, S. Ramaswamy, Mesoscopic theory for fluctuating active nematics. *New J. Phys.* **15**, 085032 (2013)
22. I.S. Aranson, L. Kramer, The world of the complex Ginzburg-Landau equation. *Rev. Mod. Phys.* **74**, 99 (2002)
23. F. Peruani, J. Starruss, V. Jakovljevic, L. Søggaard-Andersen, A. Deutsch, M. Bär, Collective motion and nonequilibrium cluster formation in colonies of gliding bacteria. *Phys. Rev. Lett.* **108**, 098102 (2012)

Chapter 4

Beyond Assemblies of Stable Units

In the previous section, we have focused on assemblies of interacting units. Such units may correspond to very different underlying basic systems, ranging from passive or active particles to social agents as in the Schelling model. In spite of these differences, all these units shared the important property of being stable, in the sense that each of them has specific given properties, that remain valid at any time. They thus constitute permanent building blocks of the system. In the present chapter, we aim at going one step further by considering situations where the elementary units are not necessarily stable in the above sense. This includes cases where particles may be created or annihilated like in reaction-diffusion processes (Sect. 4.1), as well as situations where the individual properties evolve on long time scales (Sect. 4.2), like in biological evolution (which also incorporates birth and death processes). Besides, instead of considering the particles themselves, one can consider the graph of their interactions. This is a complex, dynamically evolving object, whose links are not conserved but can be added and deleted in a random way. This motivates the presentation of some basic results on random graphs in Sect. 4.3, focusing on static aspects.

4.1 Non-conserved Particles: Reaction-Diffusion Processes

Reaction-diffusion processes are simplified models describing the evolution with time of an assembly of different types of molecules that diffuse and chemically react upon encounter. Particle types are usually described by letters, 'A', 'B', 'C', etc. Transition rates in reaction-diffusion models are often written in terms of chemical reactions, like $A + B \rightarrow C$, or $2A \rightarrow \emptyset$. Many reaction-diffusion models are defined on a lattice, in such a way that all particles sit on a node of the lattice. A microscopic configuration of the system is then given by the list of the numbers n_i^A, n_i^B, n_i^C , etc. of particles A, B, C, \dots on node i . In some models, an exclusion principle is present, so that at most one particle can lie on a given site. In this case, it is convenient to

represent the configuration of the system by introducing a local variable $q_i = 0, 1, 2, 3, \dots$ corresponding respectively to having zero particle, a particle A , a particle B , or a particle C on site i .

In the following, we consider reaction-diffusion models with a single type of particles, denoted as A , and we discuss two different types of mean-field approaches. The first one is the most common one, which simply consists in writing non-linear evolution equations for the density field ρ of particles A . This approach is a mean-field one in the sense that some correlations between particles are neglected. In addition, such an approach describes local average values, and does not account for density fluctuations. To deal with fluctuations, a second approach is to consider a fully-connected model, in which any particle can interact with any other particle, and to determine the distribution of the number of particles in the system. Such an approach, accounting for fluctuations, however loses spatial information. It effectively amounts to working in an infinite-dimensional space, similarly to the fully-connected Ising model.

4.1.1 Mean-Field Approach of Absorbing Phase Transitions

For definiteness, we will consider a simple example of reaction-diffusion model, described by the following three reactions:



The precise meaning of the rates κ , ν and λ will appear below. We further assume that particles A diffuse in space with a diffusion coefficient D . We now wish to determine an evolution equation for the density field $\rho(\mathbf{r}, t)$ describing the average number of particles in a small volume around point \mathbf{r} . The rate of change of the density resulting from reaction (4.1) is simply given by $\kappa\rho$, since for each particle already present in the system, a new particle is created with probability κ per unit time. Similarly, the rate of change associated to reaction (4.2) is equal to $-\nu\rho$, corresponding to a decrease of density with rate ν . These contributions to the evolution of ρ do not involve any approximation, being linear terms. Approximations become necessary when dealing with interactions between particles. This is the case for reaction (4.3), which involves the encounter of two particles. Strictly speaking, the probability to find two particles at the same point \mathbf{r} is a quantity that depends on correlations between the positions of particles, and that can thus not be expressed in a direct way as a function of the density field. As a first approximation, one can however neglect correlations and simply express the probability to find two particles in \mathbf{r} at time t as the square of the density $\rho(\mathbf{r}, t)$. The rate of change of the density resulting from reaction (4.3) thus simply reads $-\lambda\rho^2$. Taking also into account the diffusion of particles, we end up with the following evolution equation for the density field,

$$\frac{\partial \rho}{\partial t} = (\kappa - \nu)\rho - \lambda \rho^2 + D\Delta \rho \quad (4.4)$$

where Δ is the Laplacian operator.¹ The diffusion term tends to smooth out spatial heterogeneities of the density field. In the limit of a uniform field, Eq. (4.4) reduces to

$$\frac{\partial \rho}{\partial t} = (\kappa - \nu)\rho - \lambda \rho^2. \quad (4.5)$$

When $\kappa < \nu$, the state $\rho = 0$ is the only stationary state, and it is linearly stable as can be checked easily by linearizing Eq. (4.5) around $\rho = 0$. In contrast, for $\kappa > \nu$, the state $\rho = 0$ becomes unstable, and a new stationary state emerges, given by

$$\rho_0 = \frac{\kappa - \nu}{\lambda} \quad (\kappa > \nu). \quad (4.6)$$

A straightforward stability analysis shows that this state ρ_0 is linearly stable. The transition occurring at $\kappa = \nu$ is called an absorbing phase transition. The state $\rho = 0$ is denoted as the absorbing phase for $\kappa < \nu$, while the phase ρ_0 , present for $\kappa > \nu$, is called the active phase. Absorbing phase transitions constitute one of the major types of out-of-equilibrium phase transitions. Similarly to equilibrium phase transitions, they are characterized by a diverging correlation length ξ , and by a set of critical exponents, leading to the identification of universality classes. An important difference with respect to equilibrium phenomena is the role played by time, since a detailed characterization of absorbing phase transitions involves space-time trajectories, leading to the introduction of a correlation time τ . For a d -dimensional system, the phase transition is thus characterized as a $(d + 1)$ -dimensional process. Denoting as ε the control parameter of the transition ($\varepsilon \equiv \kappa - \nu$ in Eq. (4.4)), the stationary density ρ_0 in the active phase $\varepsilon > 0$ scales as $\rho_0 \sim \varepsilon^\beta$, which defines the exponent β . Two other critical exponents are associated to the correlation length and time, which respectively scale as $\xi \sim |\varepsilon|^{-\nu_\perp}$ and $\tau \sim |\varepsilon|^{-\nu_\parallel}$. Notations ν_\perp and ν_\parallel are standard, and come from the geometrical interpretation of the spatio-temporal process in a $(d + 1)$ -dimensional space. Universality classes are determined (in the simplest cases) by the set of critical exponents $(\beta, \nu_\perp, \nu_\parallel)$. Note that similarly to equilibrium phase transitions, the exponents characterizing the critical divergence of length and time correlations are equal above and below the transition.

The prominent universality class for absorbing phase transitions is called Directed Percolation, often abbreviated as DP. Other universality classes also exist, for instance for systems with conservation laws [1]. The reaction-diffusion process described by Eqs. (4.1)–(4.3) is a typical example of a process belonging to the DP universality class [1]. Equation (4.4) is thus a mean-field representation of the DP class. One sees from Eq. (4.6) that the mean-field value of the β exponent for the DP class is $\beta^{\text{MF}} = 1$. The mean-field value of the exponent ν_\parallel can also be easily

¹The Laplacian operator is defined as $\Delta = \partial^2/\partial x^2$ in one dimension, $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ in two dimensions, and $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ in three dimensions.

determined from Eq. (4.5). In the inactive phase $\varepsilon \equiv \kappa - \nu < 0$, the density decays to zero and is thus asymptotically described by the linearized equation $\partial\rho/\partial t = \varepsilon\rho$ (with $\varepsilon < 0$), whose solution is

$$\rho(t) \propto e^{-|\varepsilon|t}. \quad (4.7)$$

As the correlation time is defined by $\rho(t) \propto e^{-t/\tau}$, it follows from Eq. (4.7) that $\tau = |\varepsilon|^{-1}$, resulting in a mean-field exponent $\nu_{\parallel}^{\text{MF}} = 1$. Finally, scaling arguments performed on Eq. (4.4) allows the mean-field value of the exponent ν_{\perp} to be determined as well, yielding $\nu_{\perp}^{\text{MF}} = \frac{1}{2}$ [1]. Using field-theoretical arguments, one can show that these mean-field exponents are correct in space dimension $d \geq d_c = 4$, where d_c is the upper critical dimension. For $d < 4$, the exponents β , ν_{\perp} and ν_{\parallel} differ from their mean-field values.

4.1.2 Fluctuations in a Fully Connected Model

The description of reaction-diffusion processes through Eq. (4.4) is purely deterministic, and provides no information about fluctuations, for instance the fluctuations of the total number of particles in the system. To go beyond this purely deterministic description, we now study the Markov process associated to the reaction rules (4.1)–(4.3) in a fully-connected geometry, meaning that any pair of particles in the system has at any time the same probability to react. In this very simplified setting, the only stochastic variable in the system is thus the total number n of particles (by contrast, a density field $\rho(\mathbf{r}, t)$ is used in the above deterministic mean-field description). The stochastic evolution of the number n of particles under the reactions (4.1)–(4.3) is described by the following transition rates:

$$W(n \rightarrow n + 1) = \kappa n \quad (4.8)$$

$$W(n \rightarrow n - 1) = \nu n + \frac{\lambda}{V} n(n - 1). \quad (4.9)$$

Note that in this fully-connected geometry, one needs to rescale the rate λ by the volume V of the system in order to obtain a well-defined infinite volume limit. The probability $P_n(t)$ to have a number n of particles at time t then obeys the master equation

$$\begin{aligned} \frac{dP_n}{dt} = & - \left((\kappa + \nu)n + \frac{\lambda}{V} n(n - 1) \right) P_n + \kappa(n - 1)P_{n-1} \\ & + \left(\nu(n + 1) + \frac{\lambda}{V} n(n + 1) \right) P_{n+1}, \end{aligned} \quad (4.10)$$

with the convention that $P_{-1} = 0$. Let us first note that the absorbing state P_n^{as} defined by $P_0^{\text{as}} = 1$ and $P_n^{\text{as}} = 0$ for $n \geq 1$ is a solution of Eq. (4.10) for all values

of the parameters. This solution is stable by definition of the absorbing state: once all particles have been annihilated, there is no way to create new particles and thus to change state. This result seems to contradict the mean-field result of Sect. 4.1.1, according to which an active phase is present for $\kappa > \nu$. The reason for this is that the mean-field approach neglects fluctuations, and that the absorbing state may be reached when $\kappa > \nu$ only through atypically large fluctuations, as shown below. This suggests that the active phase is actually a long-lived metastable state, with as we will see, a lifetime that diverges exponentially with the volume of the system.

The metastable active state can be approximately described by a probability P_n^{ms} such that $P_0^{\text{ms}} = 0$, thus excluding the state with no particle. From Eq. (4.10), we have that

$$\frac{dP_0}{dt} = \nu P_1. \quad (4.11)$$

Hence the assumption of a metastable state is consistent only if P_1^{ms} is very small, so that P_0 remains close to zero for a very long time.

In the large volume limit, we assume that the probability distribution P_n^{ms} takes a large deviation form [2] (see Sect. 6.3 for more details):

$$P_n^{\text{ms}} \propto e^{-V\phi(\rho)} \quad (n > 0), \quad (4.12)$$

where $\rho = n/V$ is the (fluctuating) density, and $\phi(\rho)$ is a large deviation function. Note that due to the normalization of the probability density, one has $\phi(\rho) \geq 0$ for all ρ . In order to use the large deviation form of P_n^{ms} in Eq. (4.10), we need to determine $P_{n\pm 1}^{\text{ms}}$, which can be done by a first order expansion of ϕ , namely

$$P_{n\pm 1}^{\text{ms}} \propto e^{-V\phi(\rho \pm \frac{1}{V})} \approx e^{-V\phi(\rho)} e^{\mp\phi'(\rho)}. \quad (4.13)$$

Expressing n as a function of ρ everywhere in Eq. (4.10), one finds in the limit $V \rightarrow \infty$, after rearranging the terms

$$(\nu\rho + \lambda\rho^2) e^{-2\phi'(\rho)} - [(\kappa + \nu)\rho + \lambda\rho^2] e^{-\phi'(\rho)} + \kappa\rho = 0. \quad (4.14)$$

The fact that a well-defined equation is obtained in the limit $V \rightarrow \infty$ shows that the assumption of a large deviation form for the distribution P_n^{ms} is consistent. The quadratic equation (4.14) for the variable $e^{-\phi'(\rho)}$ can be solved, yielding two potential solutions, the relevant one being

$$e^{-\phi'(\rho)} = \frac{\kappa}{\nu + \lambda\rho}. \quad (4.15)$$

We thus end up with

$$\phi'(\rho) = \ln \frac{\nu + \lambda\rho}{\kappa}. \quad (4.16)$$

We already see from this equation that for $\kappa > \nu$, the solution ρ_0 found in Eq. (4.6) yields $\phi'(\rho_0) = 0$, and thus corresponds to the most probable value of ρ , consistently with the deterministic mean-field picture developed in Sect. 4.1.1. That ρ_0 is indeed a minimum of ϕ (a maximum of the probability distribution) can be checked explicitly using the results below. Integrating $\phi'(\rho)$ in Eq. (4.16) and choosing the integration constant such that $\phi(\rho_0) = 0$, we get

$$\phi(\rho) = -(\rho - \rho_0)(1 + \ln \kappa) + \frac{1}{\lambda}(\nu + \lambda\rho) \ln(\nu + \lambda\rho) - \frac{\kappa}{\lambda} \ln \kappa. \quad (4.17)$$

The behavior of the function $\phi(\rho)$ is illustrated in Fig. 4.1.

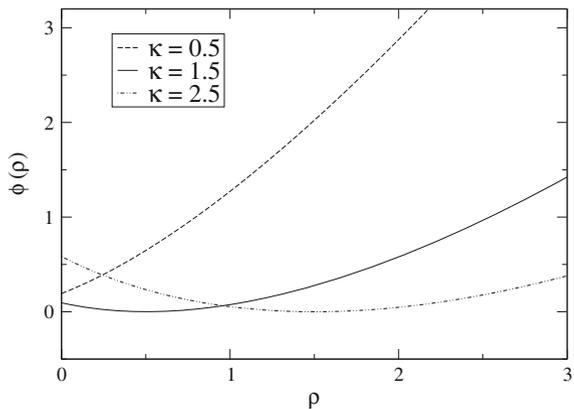
As mentioned above, for $\kappa > \nu$ the function $\phi(\rho)$ has a minimum (equal to zero) for $\rho = \rho_0$. It follows that $\phi(0) > 0$, so that $P_1^{\text{ms}} \approx e^{-V\phi(0)}$ is exponentially small with the volume. The assumption of a metastable active state is thus consistent, according to Eq. (4.11). The lifetime τ_{ms} of the metastable state can be estimated as the inverse of the rate of increase of P_0 , leading to

$$\tau_{\text{ms}} \approx \frac{1}{\nu} e^{V\phi(0)}. \quad (4.18)$$

The lifetime is thus found to increase exponentially with the volume of the system, which justifies the fact to consider the active phase as a ‘true’ phase of the system in the large size limit. In contrast, for $\kappa < \nu$, the function $\phi(\rho)$ has no minimum for $\rho > 0$, but is minimum for $\rho = 0$. As a result, $\phi(0) = 0$ so that P_1^{ms} remains of the order of 1 even for large volume V . Hence P_0 increases rapidly according to Eq. (4.11), and the system converges to the absorbing state in a relatively short time. The assumption of a metastable active state is thus no longer consistent in this case. One thus recovers the absorbing phase transition occurring at $\kappa = \nu$.

One of the interests of the present large deviation approach is to be able to describe the fluctuations of the number of particles in the active phase. Expanding $\phi(\rho)$ around

Fig. 4.1 Illustration of the large deviation function $\phi(\rho)$ for different values of κ ($\nu = \lambda = 1$). The active phase ($\kappa > 1$) corresponds to the situation where the minimum of $\phi(\rho)$ is reached for $\rho > 0$



its maximum to second order in $\rho - \rho_0$, we can compute the variance $\sigma_\rho^2 = \langle (\rho - \rho_0)^2 \rangle$ in the Gaussian approximation, yielding for the relative standard deviation

$$\frac{\sigma_\rho}{\rho_0} = \frac{\lambda}{\kappa - \nu} \sqrt{\frac{\kappa}{\lambda V}}. \quad (4.19)$$

Hence the relative amplitude of fluctuations is inversely proportional to the square root of the volume, as expected from the Central Limit Theorem (see Sect. 6.1.1). One may also note from Eq. (4.19) that the relative amplitude of density fluctuations diverges when $\kappa - \nu \rightarrow 0$.

To conclude on this model, it is also of interest to briefly comment on the analogies and differences with the fluctuations of the number of particles in the equilibrium grand canonical ensemble (see Sect. 1.3.3), in which particles are exchanged with a reservoir. Assuming for instance that particles are randomly distributed among a large number of boxes, the entropy reads $S(\rho) = -V\rho \ln \rho$, where V is here the number of boxes. Taking into account the contribution of the chemical potential μ characterizing the particle reservoir, one finds a large deviation function

$$\phi_{\text{GC}}(\rho) = \rho \ln \rho - \mu\rho + \phi_0, \quad (4.20)$$

where the constant ϕ_0 is chosen such that $\phi(\bar{\rho}) = 0$, $\bar{\rho}$ being the average density. Note that we have set the temperature to $T = 1$ as it plays no role here. In spite of some similarities, it is thus not possible to directly map the nonequilibrium large deviation function $\phi(\rho)$ given in Eq. (4.17) to an equilibrium one as given in Eq. (4.20), showing again that an absorbing phase transition is a genuine nonequilibrium phenomenon.

4.2 Evolutionary Dynamics

4.2.1 Statistical Physics Modeling of Evolution in Biology

Another example of a system composed of a large but non conserved number of units is the case of biological populations (of bacteria for instance), that evolve generation after generation under the combined effect of selection and random mutations. This is of course a topic of very general interest and with a broad literature, most of it being outside physics journals. Giving even a brief summary of what has been done in this field is clearly not possible in a few pages. The reader interested in this field is referred to specific reviews like Ref. [3].

Our more modest goal here is to give a flavour of the important similarities and differences between the dynamics of an assembly of physical units, and the evolution dynamics of a biological population. On the shortest time scales we are considering here, a biological population evolves through the birth and death of individuals. The number of individuals is thus not conserved, similarly to the reaction-diffusion

processes we have described in Sect. 4.1. However, the focus of studies of biological evolution is not on the statistics of the number of individuals in the population but rather on the statistics of the characteristics of these individuals. Contrary to the simple particles considered in reaction-diffusion processes that carry no information, individuals in a biological population are characterized, even in the simplest models, by a ‘genome’ that encodes their genetic information. This genetic information has two main roles in the models. First, it determines the fitness of an individual, that is its relative capacity to have offsprings (‘children’) as compared to other individual in the population. The higher its fitness, the more offsprings an individual is statistically expected to have: this is the selection process. The fitness characterizes the degree of adaptation of an individual to its environment. Second, the genetic information is transmitted to the offsprings, who thus carry the same genetic information as their ancestor, up to rare mutations (‘errors’ in copying the genetic code). Such mutations are also essential for the long-term evolution of the population and its adaptation to changes in the environment.

Since the primary interest is not on fluctuations of population size but on statistics of genetic information, a usual trick is to use the following type of dynamics. Considering a population of N individuals, an individual is randomly chosen with a probability per unit time proportional to its fitness. Once chosen, it gives birth to an offspring which replaces another individual, randomly chosen with uniform probability. With some small probability, the offspring may also be affected by a mutation. Such types of rules hence (somehow artificially) ensure a constant population size, making analytical treatments easier.

Forgetting about the underlying individuals in the model, one can map the above model of birth and death onto a model of N genomes subjected to a jump dynamics: the genome of the new-born individual simply replaces that of the dead individual. Note that this change of genome should not be confused with mutations. Mutations corresponds to (often small) changes in the genome of an offspring with respect to its direct ancestor. In contrast, we are here formally replacing a genome, on the list of genomes of the population, by the genome of another individual, without any filiation between the two underlying individuals.

For concreteness, let us denote as σ_i the genome of individual i . In practice, σ_i is generically an ordered list $\sigma_i = (s_{i,1}, \dots, s_{i,L})$ of L symbols belonging to a finite alphabet, for instance $s_{i,j} \in \{A, U, G, C\}$ like in DNA, or $s_{i,j} \in \{0, 1\}$ in simplified models. We will, however, not refer to the detailed structure of the variables σ_i in the following, but simply use the assumption that σ_i takes a finite number of discrete values.

Given the stochastic nature of the birth-death process described above, standard methods of nonequilibrium statistical physics suggest to describe the dynamics of the population with a master equation for the probability $P(C, t)$ that the population has a configuration $C = (\sigma_1, \dots, \sigma_N)$, characterizing the list of genomes of all the individuals in the population. Such an approach however leads to a complicated master equation that cannot be solved easily, and is thus not very helpful in practice.

An alternative solution is to describe the stochastic dynamics at the level of an individual genome σ , instead of considering explicitly the full population. Yet, due

to the presence of effective interactions between genomes generated by the fitness, transition rates between different values of σ are not predefined, time-independent functions. This results in a non-linear master equation instead of the standard linear one, as explained below. A further difficulty is to be able to determine the evolution of the probability distribution $P(\sigma, t)$ under the combined effect of birth-death processes and mutations. A standard way out of this difficulty is to assume that the mutation rate is very low, and to take into account a separation of time scales between the relatively fast birth-death process, and the much slower dynamics of mutations. We thus first consider the convergence to a steady state for the dynamics in the absence of mutations, and in a second step describe the quasistatic evolution of this steady-state under a very low mutation rate.

4.2.2 Selection Dynamics Without Mutation

We start by considering the dynamics of a population of fixed size N under the selection birth-death process without mutations; this model is a simple version of the so-called Moran model. The fitness is assumed to depend only on the genome σ , and is denoted as $f(\sigma)$. The continuous time dynamics is defined as follows. An individual with genome σ' is randomly chosen with a probability per unit time proportional to its fitness $f(\sigma')$. The chosen individual then gives birth to an offspring having exactly the same genome σ' , due to the absence of mutations. This offspring replaces another individual with genome σ , randomly chosen among the population with uniform probability. Our goal is to write down an effective master equation for the probability distribution $P(\sigma, t)$ of genomes at time t . We first need to determine the transition rate $W(\sigma'|\sigma)$ from a genome σ to a genome σ' . The process is driven by the choice of the individual with genome σ' who gives birth to an offspring, so that the transition rate does not depend on the genome σ of the replaced individual. The transition rate $W(\sigma'|\sigma)$ is simply proportional to the fitness $f(\sigma')$ and to the number $n(\sigma', t)$ of individuals with genome σ' at time t in the population. We thus end up with

$$W_t(\sigma'|\sigma) = f(\sigma') \frac{n(\sigma', t)}{N} \quad (4.21)$$

where the factor $1/N$ has been included to normalize the time scale (a frequency unit could be included here, but we have assumed the transition rates to be dimensionless). Note that we have made explicit in the notation the time-dependence of the transition rate. In this form, the transition rate $W_t(\sigma'|\sigma)$ is however not well-defined, due to the fact that $n_t(\sigma')$ is actually a random variable, depending on the global configuration of the population. This problem is nevertheless solved in the infinite population size limit $N \rightarrow \infty$, for which Eq. (4.21) reduces to

$$W_t(\sigma'|\sigma) = f(\sigma') P(\sigma', t). \quad (4.22)$$

Now $W_t(\sigma'|\sigma)$ is no longer a random variable, but is determined in a self-consistent way by the solution of the master equation

$$\frac{\partial P}{\partial t}(\sigma, t) = \sum_{\sigma' (\neq \sigma)} [W_t(\sigma|\sigma')P(\sigma', t) - W_t(\sigma'|\sigma)P(\sigma, t)], \quad (4.23)$$

the price to pay being that this master equation is now non-linear. In order to find the steady-state solution of Eq. (4.23), it is natural to first look at the simplest type of solution, namely solutions satisfying detailed balance (see Sect. 2.1.2). In the present case, detailed balance corresponds to the following equality (we drop the explicit time dependence of the transition rates since we are considering a steady state solution)

$$W(\sigma|\sigma')P(\sigma') = W(\sigma'|\sigma)P(\sigma). \quad (4.24)$$

Taking into account the expression (4.22) of the transition rates, Eq. (4.24) leads to the self-consistent solution

$$P(\sigma) = \frac{1}{Z} f(\sigma)P(\sigma) \quad (4.25)$$

where Z is a normalization factor. This equation implies that $P(\sigma)$ is non-zero only over the configurations σ having a common value f_0 of the fitness $f(\sigma)$. In the following, we assume for simplicity that all genomes σ have distinct fitnesses $f(\sigma)$. Under this assumption, the stationary probability distribution $P(\sigma)$ concentrates on a single genome σ_0 ,

$$P(\sigma) = \delta_{\sigma, \sigma_0}. \quad (4.26)$$

This phenomenon, by which the whole population acquires the same genome, is called fixation of the genome σ_0 .

A dynamical view on the fixation phenomenon can be obtained by computing the rate of change of the mean fitness

$$\langle f \rangle = \sum_{\sigma} f(\sigma)P(\sigma, t). \quad (4.27)$$

Starting from Eqs. (4.22) and (4.23), one has

$$\frac{d}{dt}\langle f \rangle = \sum_{\sigma, \sigma'} f(\sigma)[f(\sigma) - f(\sigma')]P(\sigma, t)P(\sigma', t). \quad (4.28)$$

Note that we have dropped the constraint $\sigma \neq \sigma'$ in the sum, since the corresponding term is equal to zero. Expanding the term between brackets in Eq. (4.28) directly yields

$$\frac{d}{dt}\langle f \rangle = \langle f^2 \rangle - \langle f \rangle^2. \quad (4.29)$$

Hence in the absence of mutation, the rate of change of the mean fitness in an infinite population is equal to the variance of the fitness across the population. This important result is known as Fisher's theorem of natural selection. It shows in particular that the mean fitness of the population cannot decrease, and remains constant only when the variance of the fitness vanishes, corresponding to a population in which all individuals have the same fitness. If all genomes have different fitnesses, a single genome is thus selected in the long time limit, and one recovers the phenomenon of fixation described above.

A natural question is then to know what is the probability of fixation of a given genome σ_0 starting from a heterogeneous population in which many different genomes are present. Obviously, a genome with a high fitness $f(\sigma_0)$ should have a higher probability of fixation than a genome with a low fitness. Giving a quantitative answer for an arbitrary initial condition is however a complicated problem. In the following subsection, we will discuss a specific case in which a relatively simple answer can be given.

4.2.3 Quasistatic Evolution Under Mutation

We have seen above that an initial population with heterogeneous genomes and fitnesses across individuals evolves under a birth-death process with selection (but no mutation) to a homogeneous population where all individuals have the same genome. Under the time scale separation hypothesis mentioned above, one can assume that the typical time between mutations is much larger than the time needed for the population to relax to a single genome. In this framework, the effect of successful mutations (those that reach fixation) can thus be conceived simply as jumps between different values of the genome—while mutations that do not reach fixation can simply be neglected. Since mutations are random, one needs to resort to a stochastic description, involving transition rates from one value of the genome to another. We denote as $Q(\sigma)$ the distribution of the genome σ . Let us emphasize the different interpretation of the distribution $Q(\sigma)$ with respect to the distribution $P(\sigma)$ introduced in Sect. 4.2.2. The function $P(\sigma)$ describes the distribution of genomes across the population, in the limit of a large population size where the fluctuations of this distribution can be neglected. By contrast, the distribution $Q(\sigma)$ describes a large ensemble of populations, each population being homogeneous with a single genome σ .

In order to describe the dynamics of the distribution $Q(\sigma)$, we need to determine the probability per unit time of the fixation of a mutated genome. This probability per unit time is the product of the mutation rate μ (the probability per unit time that a random mutation occurs) and of the probability that the new genome eventually reaches fixation. This probability of fixation can be evaluated as follows. Let us consider an initially homogeneous population of genome σ_1 . A random mutation then replaces one of the N genomes σ_1 by a new genome σ_2 . For brevity, the associated fitnesses $f(\sigma_1)$ and $f(\sigma_2)$ are denoted as f_1 and f_2 respectively. Under the population dynamics, the number n of genomes σ_2 in the population evolves in a stochastic

manner. When an offspring replaces another individual, several situations may occur: (i) the offspring has genome σ_2 and replaces a genome σ_1 , in which case n increases by one; (ii) the offspring has genome σ_1 and replaces a genome σ_2 , in which case n decreases by one; (iii) the genome of the offspring is the same as the replaced one, so that n does not change. In case (i), the probability that the offspring has genome σ_2 is proportional to $f_2 n$ (each of the n genomes σ_2 has a probability proportional to f_2 to be chosen in order to give an offspring), while the probability that the randomly replaced individual has genome σ_1 is proportional to $N - n$ (it is picked up in a uniform way among the $N - n$ genomes σ_1). The probability $T(n + 1|n)$ to increase n by one is then given by

$$T(n + 1|n) = C_n f_2 n (N - n) \quad (4.30)$$

where C_n is a normalization constant. Similarly, in case (ii), the probability that the offspring has genome σ_1 is proportional to $f_1 (N - n)$, while the probability that the randomly replaced individual has genome σ_2 is proportional to n . The probability $T(n - 1|n)$ that n decreases by one then reads

$$T(n - 1|n) = C_n f_1 n (N - n). \quad (4.31)$$

The normalization constant C_n is determined by the condition that $T(n + 1|n) + T(n - 1|n) = 1$, yielding

$$T(n + 1|n) = \frac{f_2}{f_1 + f_2}, \quad T(n - 1|n) = \frac{f_1}{f_1 + f_2}. \quad (4.32)$$

The problem thus boils down to a random walk for the integer n in the interval $0 \leq n \leq N$, with transition rates given by Eq. (4.32). The walk stops when it reaches either $n = 0$ or $n = N$. For brevity, we introduce the notations $q \equiv T(n + 1|n)$; hence $1 - q = T(n - 1|n)$. With probability one, the walk eventually reaches either $n = 0$ (in which case the mutated genome σ_2 disappears from the population) or $n = N$ (in which case the genome σ_2 is fixed). The probability that the walk does not reach any boundary after an infinite number of steps is zero. The fixation probability is the probability for the walk to reach $n = N$ starting from $n = 1$. To compute the fixation probability, let us introduce more generally the probability $P_f(m)$ that the walk first reaches $n = N$ starting from a position $n = m$. The fixation probability is then simply $P_f(1)$. The interest of introducing $P_f(m)$ lies in the fact that this quantity obeys a recursion relation, namely

$$P_f(m) = q P_f(m + 1) + (1 - q) P_f(m - 1) \quad (4.33)$$

with $1 \leq m \leq N - 1$. The interpretation of this relation is very simple. Starting from position $n = m$, the walk can either jump to $m + 1$ with probability q and then have a probability $P_f(m + 1)$ to eventually reach $n = N$, or jump to $m - 1$ with probability $1 - q$ and have a probability $P_f(m - 1)$ to reach N . Equation (4.33) can be rewritten

as

$$P_f(m+1) - P_f(m) = r[P_f(m) - P_f(m-1)] \quad (4.34)$$

with

$$r = \frac{1-q}{q} = \frac{f_1}{f_2} \quad (4.35)$$

where we have taken into account the definition Eq. (4.32) of the transition rates. By summation, $P_f(m)$ can be obtained from Eq. (4.34) as

$$P_f(m) = P_f(0) + [P_f(1) - P_f(0)] \sum_{k=0}^{m-1} r^k \quad (4.36)$$

for $m = 1, \dots, N$. Computing explicitly the geometric sum, and taking into account the boundary conditions $P_f(0) = 0$ and $P_f(N) = 1$, one obtains

$$P_f(m) = \frac{1-r^m}{1-r^N} \quad (4.37)$$

from which the fixation probability follows,

$$P_f(1) = \frac{1-r}{1-r^N}. \quad (4.38)$$

Note that although we may have in mind a beneficial mutation, that is $f_2 > f_1$, the fixation probability (4.38) is valid whatever the values of f_1 and f_2 . With the above notation $r = f_1/f_2$, one has for large N that $P_f(1) \approx (r-1)$ if $f_2 > f_1$, and $P_f(1) \approx (1-r)/r^N \ll 1$ if $f_2 < f_1$. Hence a beneficial mutation ($f_2 > f_1$) has a finite probability of fixation, while a deleterious mutation ($f_2 < f_1$) has a very small fixation probability, that goes to zero when the population size N goes to infinity.

Considering the long time scale dynamics through which the genome of the homogeneous population changes under mutation and fixation of the mutated genome, the transition rate from σ_1 to σ_2 reads

$$W_m(\sigma_2|\sigma_1) = \mu \frac{1 - (f_1/f_2)}{1 - (f_1/f_2)^N} \quad (4.39)$$

with μ the mutation rate. These transition rates govern the evolution of the distribution $Q(\sigma, t)$ according to

$$\frac{\partial Q}{\partial t}(\sigma, t) = \sum_{\sigma' (\neq \sigma)} [W_m(\sigma|\sigma') Q(\sigma', t) - W_m(\sigma'|\sigma) Q(\sigma, t)]. \quad (4.40)$$

One can easily check that the transition rate (4.39) satisfies a detailed balance relation of the form [4]

$$W_m(\sigma_2|\sigma_1) f_1^{N-1} = W_m(\sigma_1|\sigma_2) f_2^{N-1}. \quad (4.41)$$

As a result, the stationary probability distribution $Q(\sigma)$, reached in the infinite time limit, is proportional to $f(\sigma)^{N-1}$,

$$Q(\sigma) = \frac{1}{Z} f(\sigma)^{N-1} \quad (4.42)$$

where Z is a normalization factor. Let us emphasize that N is simply an external parameter in this long time scale dynamics of the genome σ , since the population is homogeneous and the definition of σ does not involve N . This is to be contrasted for instance with the full configuration $(\sigma_1, \dots, \sigma_N)$ of an heterogeneous population, which depends on N . Equation (4.42) suggests an interesting analogy with equilibrium statistical physics [4]. Defining $\varepsilon(\sigma) = -\ln f(\sigma)$, one can rewrite Eq. (4.42) as

$$Q(\sigma) = \frac{1}{Z} e^{-\beta_{\text{eff}}\varepsilon(\sigma)}. \quad (4.43)$$

This form of the distribution $Q(\sigma)$ shows a clear analogy to the equilibrium distribution in statistical physics, provided one interprets $\varepsilon(\sigma)$ as an effective energy, and $\beta_{\text{eff}} = N - 1$ as an effective inverse temperature. Having an effective temperature which depends on N may be surprising at first sight, but let us emphasize again that N can here be considered as an external parameter, as explained above. This property has important consequences. In the infinite size limit, the effective temperature $T_{\text{eff}} = \beta_{\text{eff}}^{-1}$ is equal to zero, and the distribution concentrates on the lowest energy states, that is on the genomes with the highest fitness. For a finite population size, fluctuations around these states of highest fitness are allowed, and these fluctuations become larger when population size is decreased. By analogy with the potential energy landscape of physical systems, it is customary to speak about the “fitness landscape” in the context of biological evolution modeling. Note that, strictly speaking, the fitness landscape characterizes a single genome; it is simply a representation of the value of the fitness as a function of the genome value. However, the notion of fitness landscape is more useful to describe a population, especially under the simplifying assumption that the population is homogeneous and can be described by a single genome. It is only by considering a population that a dynamics in the fitness landscape can be defined, and we have seen that the population size N plays a key role as being essentially the inverse effective temperature. In a complex fitness landscape, a small population size (relatively high effective temperature) may help to reach higher values of the fitness by escaping local maxima in the fitness landscape thanks to fluctuations that may temporarily decrease the fitness.

As mentioned earlier, the genome σ is often described in evolution models as a sequence of symbols, $\sigma = (s_1, \dots, s_L)$, with in the simplest models $s_j \in \{0, 1\}$ or $s_j \in \{-1, 1\}$. The integer L is in general assumed to be large. In the context of these models, there is thus a natural mapping between the fitness landscape and the energy landscape of spin models in physics. In particular, mappings to disordered

spin models like spin glass models have been proposed [3]. Typical examples of fitness functions directly inspired by disordered spin models include the analogue of the random field paramagnetic model

$$f(\sigma) = \sum_{i=1}^L h_i s_i + F_0, \quad \sigma = (s_1, \dots, s_L), \quad (4.44)$$

where h_i is a quenched random variable, as well as the p -spin model

$$f(\sigma) = \sum_{i_1, i_2, \dots, i_p} J_{i_1, i_2, \dots, i_p} s_{i_1} s_{i_2} \dots s_{i_p} + F_0 \quad (4.45)$$

where $p \geq 3$ is an integer parameter of the model. The constant F_0 is included to ensure that the fitness remains positive. The parameters J_{i_1, i_2, \dots, i_p} are (time-independent) random coupling constants that couple the p spins $(s_{i_1}, \dots, s_{i_p})$. The distribution of these coupling constants in general depends on the total number L of spins. The sum is performed over all sets of p spins among the L spins. The random field model (4.44) is a simple realization of a so-called ‘Fujiyama landscape’ [3], in which there is a single maximum in the landscape, that is reached under evolution from any initial genome. By contrast, the p -spin model yields a complicated fitness landscape that includes many local maxima in which the population may get trapped during evolution.

A popular alternative to the p -spin model is the so-called NK-landscape, which associates to each of the L spins σ_i a set of K (typically randomly chosen) ‘neighbors’, with $K < L$. The values of these neighbor spins determine the contribution of spin s_i to the total fitness, according to

$$f(\sigma) = \frac{1}{L} \sum_{i=1}^L \tilde{J}_i(s_i, s_{i_1}, s_{i_2}, \dots, s_{i_K}). \quad (4.46)$$

The parameters $\tilde{J}_i(s_i, s_{i_1}, s_{i_2}, \dots, s_{i_K})$ are quenched random variables that take statistically independent values for each configuration $(s_i, s_{i_1}, \dots, s_{i_p})$. One of the main differences with the p -spin model is that in the NK-model, each spin s_i interacts with a single set of K spins, considered as its neighbors. In contrast, in the p -spin model a given spin s_i interacts with all possible sets of $p - 1$ other spins.

4.3 Dynamics of Networks

In this chapter, we have up to now considered the dynamics of non-conserved particles. More generally, interacting particles often generate a complex network of interactions, which evolves in time. The study of dynamically evolving networks is

in itself a topic of interest, which can be interpreted as the creation, annihilation and rewiring of links between a set of nodes, and can thus be considered as the dynamics of interacting non-conserved units—the links. As an elementary introduction to the statistics of networks, we will here mostly focus on statistical properties of static networks. We refer the reader to more advanced reading, like Ref. [5], for an introduction to dynamical aspects of networks.

Statistical physicists, as well as condensed matter physicists, are familiar with the use of lattices to model crystals, or more generally many types of models of interacting particles or agents having interaction with other particles in a well-defined neighborhood. Typical examples include the Ising model, or the original version of the Schelling model, which is also defined on a two-dimensional lattice. The lattice geometry is relevant when the metric (or Euclidean) distance is the criterion deciding which sites are linked: a link is then present only between the closest sites, in the sense of the standard metric distance of a continuous Euclidean space.

However, this regular network geometry is not relevant in all cases, and examples where more complex (or less regular) networks are useful abound in the complex system literature. In the last two decades, this field has known a very intense research activity, partly driven by the increase of computer power (both in terms of memory and computation speed) and the somewhat related availability of large data sets coming from the web (e.g., network of hyperlinks between websites) or other sources (airport network with passenger traffic, data on the spreading of epidemics, etc.) [6]. For those types of applications, regular networks like lattices are of little help, while complex networks including some type of randomness have proven useful in the description of these real world data. More information on this topic can be found for instance in Refs. [5, 6].

4.3.1 *Random Networks*

A network (or graph) is basically defined by a set of N nodes (also called vertices, or sites), and a list of links (or edges) between these nodes. One can define a variable g_{ij} which is equal to 1 when there is a link between i and j , and equal to 0 otherwise. Links may be directed or not. A directed link means that the link is defined from i to j . Hence the fact that i is linked to j does not imply that j is linked to i ; to do so, a second link has to be drawn from j to i . If links are directed, the network is called a directed graph. From the knowledge of lattices in statistical physics, we are more familiar with undirected graphs. For undirected graphs, the matrix formed by the coefficients g_{ij} is symmetric by construction. For a directed graph, the matrix g_{ij} is not necessarily symmetric, but it may also be in some limit cases, if all links between two sites come in pair with opposite orientations. This limit case is however of low interest, as it boils down to having an undirected graph. So in practice, directed graphs have asymmetric matrices g_{ij} .

For a random graph, one has to assign to each possible realization of a graph with N nodes (that is, to each matrix g_{ij}) a given probability. The simplest type of random network is the so-called Erdős-Rényi random network. Two variants of this network actually exist. The original Erdős-Rényi model consists in assigning an equal probability to all possible graphs with N nodes and M undirected links, and a zero probability to graphs having a number of links different from M . Although conceptually simple, this model is not the most convenient one for practical calculations, and one often uses instead an alternative version of the model. In this second variant, one builds a random graph by including a link between each pair (i, j) of nodes independently with a probability p . Hence the number of links is on average $\langle M \rangle = pN(N - 1)/2$, but fluctuations around this value are allowed. The difference between these two versions of the model is somewhat similar to the difference between the microcanonical and canonical ensembles introduced in equilibrium statistical physics. In this latter context, calculations are more convenient in the canonical ensemble. Similarly, calculations of graph properties are easier in the second version of the model with independent probabilities to have a link between two nodes.

In the version of the Erdős-Rényi model having independent random links on each node with probability p , the total number M of links is a random variable with a binomial distribution

$$P(M) = \frac{M_{\max}!}{M!(M_{\max} - M)!} p^M (1 - p)^{M_{\max} - M} \quad (4.47)$$

where $M_{\max} = N(N - 1)/2$ is the number of pairs of nodes. The interpretation of the binomial distribution (4.47) is simple. The probability that a link is present on a given pair of nodes is p , and the probability to have no link is $1 - p$. Hence the probability to have M links at given positions (i.e., on given pairs of nodes) and no links elsewhere is $p^M (1 - p)^{M_{\max} - M}$, due to statistical independence. The combinatorial factor in Eq. (4.47) simply counts the number of possible positions that the M links can occupy. In the large N limit, the distribution $P(M)$ takes a large deviation form

$$P(M) \approx e^{-M_{\max} \Phi(y)} \quad (0 < y < 1) \quad (4.48)$$

with $y = M/M_{\max}$ and

$$\Phi(y) = y \ln \frac{y}{p} + (1 - y) \ln \frac{1 - y}{1 - p}. \quad (4.49)$$

The fluctuations of M on a scale $\sqrt{M_{\max}} \sim N$ are described by a Gaussian distribution of mean value $\langle M \rangle = pN^2/2$ and variance $\langle (M - \langle M \rangle)^2 \rangle = pN^2/2$.

Another simple quantity of interest is the degree k of a node, defined as the number of nodes to which a given node is connected. This degree k ranges by definition between 0 and $N - 1$. The degree distribution is also binomial, like the distribution of the total number of links, and reads

$$P_d(k) = \frac{(N-1)!}{k!(N-1-k)!} p^k (1-p)^{N-1-k}. \quad (4.50)$$

This distribution is easily interpreted as follows. A given node can potentially be connected to any of the $N-1$ other nodes. The probability to have a degree k is thus given by the probability p^k to have k links, times the probability $(1-p)^{N-1-k}$ that there is no link to the $N-1-k$ remaining nodes. The combinatorial factor in Eq. (4.50) then simply counts the number of ways to choose the k nodes among $N-1$ to connect the links.

From Eq. (4.50), the average degree is equal to $\langle k \rangle = p(N-1)$ ($\approx pN$ for large N). Hence it diverges with the size of the graph, meaning that in a large graph, any node is connected to a large number of other nodes. In many applications however, one is interested in large random graphs with a fixed (relatively low) average degree. The solution is then to choose a probability p that depends on N , namely $p = z/N$, where z is a constant. The average number of neighbors is then equal to z for large N . In this case, the distribution (4.50) simplifies, for large N , to a Poisson distribution [7]

$$P_d(k) = \frac{z^k}{k!} e^{-z} \quad (4.51)$$

which is independent of N . Under the assumption $p = z/N$, the distribution $P(M)$ of the total number of nodes takes for large N a large deviation form which differs from Eq. (4.48), namely

$$P(M) \approx e^{-N\tilde{\Phi}(x)} \quad (4.52)$$

with $x = M/N$, and

$$\tilde{\Phi}(x) = x \ln \frac{2x}{z} - x + \frac{z}{2}. \quad (4.53)$$

The distribution (4.52) concentrates around the average value $M = zN/2$. Note that the function $\tilde{\Phi}(x)$ is now defined over the entire positive real axis, while the function $\Phi(y)$ introduced in Eq. (4.49) is restricted to the interval $0 < y < 1$. Fluctuations of M on a scale \sqrt{N} are still described by a Gaussian statistics, with a variance $\langle (M - \langle M \rangle)^2 \rangle = zN/2$.

4.3.2 Small-World Networks

Another important class of complex networks, which plays an important role in the modeling of real-world data, is the small-world network introduced by Watts and Strogatz [8]. The precise definition of the Watts-Strogatz model can be found for instance in [6]. Here, we only sketch the main idea which is common to networks with small-world properties. It consists in associating properties of “metric networks” (networks where only nodes closer than a given distance in the embedding Euclidean

space are linked) with properties of random networks, like the fact that the length of the shortest path (counted in number of links along the network) remains relatively small even for large networks. To do so, one basically starts from a “metric network” (e.g., a lattice) and adds to it with a small probability some random links between arbitrary (typically distant) nodes. Instead of adding links, one may also “rewire” the network, that is choose randomly a link (i, j) and replace it by a link (i, k) , where k has been chosen randomly among all nodes (possibly with some constraints). In this way, the number of links is conserved, which may be of interest in some cases.

The motivation for introducing such networks notably comes from the wish to have complex networks with a high clustering coefficient (as observed in many real-world networks), while classical Erdős-Rényi networks have a low clustering coefficient. The (local) clustering coefficient quantifies the tendency of the neighbors of a given node i_0 to be connected between themselves. If a node has k neighbors, these neighbors can have at most $n_{\max} = k(k - 1)/2$ undirected links between them (note that links to the original node i_0 are not counted). If the actual number of links between the neighbors of the node i_0 is n (again excluding links with i_0), the local clustering coefficient is given by $c(i_0) = n/n_{\max}$. An average clustering coefficient \bar{c} can be defined by averaging $c(i_0)$ over the nodes i_0 . Regular networks like lattices have a relatively large (meaning a finite fraction of unity) average clustering coefficient, while the average shortest-path length is large for a large graph. In contrast, random graphs of the Erdős-Rényi type have a low (much smaller than one) average clustering coefficient, as well as a small average shortest-path length. Yet, many real-world networks have both a relatively large clustering coefficient, and a small shortest-path length, which motivated the introduction of small-world networks satisfying this property.

As a more concrete illustration of the interest of small-world networks, let us consider the following real-world application related to transportation networks. Let us imagine that we try to analyze the railway network on the scale of a large country, or of a continent. Nodes of the network are the cities in which there is a railway station, and the links are the railways between cities. Such a network is constrained by the two-dimensional geometry of the surface of the Earth, and is thus expected to be of metric type: there are most often no direct railways between very distant cities. Now imagine that one is not interested only in the railway network, but more generally on transportation means between cities. One will thus also include in the network the airplane lines between large cities having an airport. Including these airplane lines thus drastically changes the properties of the network. With only railways, the length of the shortest path along the network typically grows as the Euclidean distance (on the surface of the Earth) between the nodes. Including the airplane lines, the ‘length’ of the shortest path (understood here as the time needed to travel along the path) grows much more slowly with distance. This is in line with our common experience. For instance, in our modern world, the time needed to travel 10000 km is much less than one hundred times the duration of a 100 km trip. Note however that this was not the case more than one century ago, when only ground transportations were available.

4.3.3 Preferential Attachment

We have seen in Eq. (4.51) that the degree distribution is a Poisson distribution, which implies that its variance is equal to the average value $\langle k \rangle$. As a result, typical fluctuations around $\langle k \rangle$ are of the order of $\langle k \rangle^{1/2}$. It is thus very unlikely to observe a node with a degree much larger than the average value. Yet, power-law distributions of degrees have been reported in many real networks [6]; such networks have been called ‘scale-free’. Even in cases where the degree distribution does not follow a power law, it still significantly differs from the Poisson distribution in most cases. A generic mechanism to account for this broader distribution has been proposed by Barabási and Albert [6]. It relies on two basic ingredients. The first one is to model the dynamics which builds the graph, instead of simply looking at the final graph. Hence the graph is built step by step, by successively adding new nodes and connecting each new node to one or several previous nodes according to some stochastic rules. The second ingredient is related to a specific property of these stochastic rules, and is called preferential attachment. The idea is that the new node tends to attach preferentially to existing nodes that already have a high degree. A simple way to do that is to assume that the probability to attach the new node to an existing node i is proportional to its degree k_i . The connectivity of the added node is fixed to a given value m , meaning that m new links are randomly attached to the existing links according to the preferential attachment rule.

Numerical simulations of this model show that the degree distribution has a power-law tail proportional to k^{-3} for large k [6]. This is again in stark contrast with the classical Erdős-Rényi random graph which has a Poisson degree distribution, which decays faster than any power-law distribution at large k . Interestingly, this power-law k^{-3} can also be predicted using relatively simple analytical arguments [6]. Let us denote as $N_k(t)$ the average number of nodes with k edges at time t . We assume that time is continuous, and that on average one node is added per unit time. According to the preferential attachment rule, the probability for each of the m links of a new node to attach at time t to a given existing node of degree k is equal to

$$q_k(t) = \frac{k}{\sum_{k'} k' N_{k'}(t)}. \quad (4.54)$$

Taking into account the fact that m links are attached to each new node, the average number $N_k(t)$ evolves according to

$$\frac{dN_k}{dt} = m(k-1)q_{k-1}(t) - mkq_k(t) + \delta_{k,m}. \quad (4.55)$$

Equation (4.55) is a balance equation formally similar to a master equation, except that the total number of nodes is not constant in time, contrary to a total probability which is constrained to remain equal to 1. The first term in Eq. (4.55) accounts for

the increase of N_k due to the attachment of a new link to a node of degree $k - 1$, which thus becomes of degree k . The second term describes the decrease of N_k due to the attachment of a new link to a node of degree k , which thus becomes of degree $k + 1$. Finally, the last term accounts for the newly added node, which has degree m and contributes only to N_m . At large time, the total number of nodes satisfies

$$\sum_k N_k(t) = t, \quad (4.56)$$

and the total number of links is equal to mt . Since a link is attached to two nodes, the average degree of the nodes is equal to $2m$, so that for large time

$$\sum_k kN_k(t) = 2mt. \quad (4.57)$$

The degree distribution $P(k)$, assumed to be time-independent in the long time regime considered here, is given by

$$P(k) = \frac{N_k(t)}{\sum_{k'} N_{k'}(t)} \quad (4.58)$$

so that $N_k(t) = tP(k)$, taking into account Eq. (4.56). In this regime, Eq. (4.55) can be rewritten as

$$P(k) = \frac{1}{2}(k - 1)P(k - 1) - \frac{1}{2}kP(k) + \delta_{k,m}. \quad (4.59)$$

At large times, all nodes have at least a degree m , since new added nodes have a degree m , and initial nodes have been connected to added nodes and have a large degree (one may also assume that the initial nodes all have a degree of at least m). We thus assume that in the large time regime, $N_k = 0$ for $k < m$. Then Eq. (4.59) reduces for $k > m$ to the following recursion relation

$$P(k) = \frac{k - 1}{k + 2} P(k - 1), \quad (4.60)$$

while the case $k = m$ provides the condition $P(m) = 2/(m + 2)$. Solving this recursion equation leads to

$$P(k) = \frac{2m(m + 1)}{k(k + 1)(k + 2)} \quad (4.61)$$

from which the large k behavior $P(k) \sim k^{-3}$ follows.

To conclude this section, we note that an interesting mapping between the random dynamics of networks and the dynamics of the Zero-Range Process has also been proposed [9]. In this case, the number of nodes is fixed, and the dynamics rather proceeds through the rewiring of links. Without entering into details, the basic idea of this mapping is that rewiring a link is similar to moving a particle from one

node to another, as if a particle was attached to the end of each link. Modeling preferential attachment actually requires a generalization of the Zero Range Process, called Misanthrope process, in which the transfer of a particle from a site to another depends on the numbers of particles on both the departure and arrival sites. With a preferential attachment dynamics, which favors rewiring to nodes with a high degree, one may obtain under some conditions a condensation transition similar to the one of the Zero Range Process (see Sect. 3.2). This condensation corresponds in the network to the onset of a hub, that is a node to which a finite fraction of all the links are attached.

Note that the above mapping is actually not exact, and requires some approximations. An exact, though more complicated mapping of a directed network to a Zero-Range process with many different types of particles has also been proposed [10].

We have considered here very basic network models, mostly undirected. Directed graphs also play an important role. Besides, to each link of the network may also be associated a ‘weight’, like the passenger traffic on a transportation line (train, plane,...); such graphs are called weighted networks. In addition to static graphs (or graph ensembles), we have briefly seen one type of dynamics which is the network growth, as in the Albert-Barabási model. This is however just a way to build the graph, but the final graph which is studied is also static. Studying the dynamics of graphs rather consists in looking at graphs which evolve dynamically, for instance due to some rewiring dynamics. Finally, recent research trends focus on the study of dynamical processes occurring on complex networks, like the spreading of an epidemic or of a rumor in a population (in which case the network represents the contacts between individuals, along which a disease or a piece of information may be transmitted). Clearly, this dynamical process may take place on a network which is itself dynamically rearranging, giving rise to an interesting and non-trivial interplay between both types of dynamics.

References

1. H. Hinrichsen, Nonequilibrium critical phenomena and phase transitions into absorbing states. *Adv. Phys.* **49**, 815 (2000)
2. H. Touchette, The large deviation approach to statistical mechanics. *Phys. Rep.* **478**, 1 (2009)
3. B. Drossel, Biological evolution and statistical physics. *Adv. Phys.* **50**, 209 (2001)
4. G. Sella, A.E. Hirsh, The application of statistical physics to evolutionary biology. *Proc. Nat. Acad. Sci. USA* **102**, 9541 (2005)
5. A. Barrat, M. Barthélémy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008)
6. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. *Rev. Modern Phys.* **74**, 47 (2002)
7. M.E.J. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118 (2001)
8. D.J. Watts, S.H. Strogatz, Collective dynamics of ‘small-world’ networks. *Nature (London)* **393**, 440 (1998)

9. M.R. Evans, T. Hanney, Nonequilibrium statistical mechanics of the zero-range process and related models. *J. Phys. A* **38**, R195 (2005)
10. A.G. Angel, T. Hanney, M.R. Evans, Condensation transitions in a model for a directed network with weighted links. *Phys. Rev. E* **73**, 016105 (2006)

Chapter 5

Statistical Description of Deterministic Systems

Although we have up to now mostly focused on stochastic descriptions of non-equilibrium systems, deterministic descriptions are also a widely used tool in complex system modeling. In the beginning of Sect. 5.1, we have already discussed some examples of deterministic dynamics, when describing for instance the time evolution of mechanical systems and some preliminary aspects of the construction of statistical physics. From Sect. 5.2 on, we have switched to a stochastic description of the systems under consideration, as this type of description turns out to be very convenient in a statistical physics context. In the present section, we come back to deterministic systems to briefly introduce some basic properties of this type of dynamics (Sect. 5.1), and to see how probabilistic tools may be of some relevance to describe chaotic deterministic systems (Sect. 5.2). In a second stage, we discuss how the coupling of a large number of dynamical systems having different parameter values may lead to a non-trivial collective behavior, like the global restabilization of unstable individual units (Sect. 5.3) or the synchronization of coupled oscillators (Sect. 5.4).

5.1 Basic Notions on Deterministic Systems

5.1.1 Fixed Points and Simple Attractors

In this first subsection, we focus on dynamical systems with continuous time dynamics. The notions introduced here can be defined in a similar way for discrete time dynamical systems. We will briefly discuss such discrete time systems in Sect. 5.1.3, when introducing the notion of chaotic dynamics.

In a continuous time description, deterministic systems are described by an ordinary differential equation

$$\frac{dx}{dt} = F(x(t)). \quad (5.1)$$

An important notion is that of fixed point of the dynamics. If a constant $x^{(0)}$ is such that $F(x^{(0)}) = 0$, then $x(t) = x^{(0)}$ is a time-independent solution of Eq. (5.1), and $x^{(0)}$ is said to be a fixed point of the dynamics. Fixed points can then be classified into two main categories according to their linear stability properties. If a value x slightly away from $x^{(0)}$ tends to converge with time to $x^{(0)}$, then $x^{(0)}$ is said to be a linearly stable fixed point. In the opposite case, the distance between $x(t)$ and $x^{(0)}$ grows with time, and $x^{(0)}$ is said to be a linearly unstable fixed point. Mathematically, one writes $x(t) = x^{(0)} + \varepsilon(t)$, assuming $\varepsilon(t)$ to be small, and linearizes Eq. (5.1), leading to

$$\frac{d\varepsilon}{dt} = F'(x^{(0)}) \varepsilon, \quad (5.2)$$

taking into account that $F(x^{(0)}) = 0$. The linear stability of the fixed point $x^{(0)}$ is simply given by the sign of $F'(x^{(0)})$: the fixed point is linearly stable for $F'(x^{(0)}) < 0$, while it is linearly unstable for $F'(x^{(0)}) > 0$. In the case where $F'(x^{(0)}) = 0$, the fixed point is said to be marginally stable, and stability is actually determined by the first nonzero term in the expansion of $F(x^{(0)} + \varepsilon)$ in powers of ε . Note also that a linearly stable fixed point may be unstable with respect to large enough perturbations. In this case, the fixed point is said to be nonlinearly unstable.

These basic notions can be easily generalized to the deterministic dynamics of several coupled degrees of freedom, described by a set of ordinary differential equations characterizing the evolution of N degrees of freedom $x_i(t)$,

$$\frac{dx_i}{dt} = F_i(x_1, \dots, x_N), \quad i = 1, \dots, N. \quad (5.3)$$

The notion of fixed point and of their stability can be introduced in the same way as above. If $(x_1^{(0)}, \dots, x_N^{(0)})$ are such that $F(x_1^{(0)}, \dots, x_N^{(0)}) = 0$, then $x_i(t) = x_i^{(0)}$ is a time-independent solution of Eq. (5.3) for all $i = 1, \dots, N$ and the point $(x_1^{(0)}, \dots, x_N^{(0)})$ is said to be a fixed point. The stability is studied by introducing a small perturbation around the fixed point:

$$x_i(t) = x_i^{(0)} + \varepsilon_i(t) \quad (5.4)$$

leading after linearization of Eq. (5.1) to

$$\frac{d\varepsilon_i}{dt} = \sum_{j=1}^N \frac{\partial F_i}{\partial x_j} (x_1^{(0)}, \dots, x_N^{(0)}) \varepsilon_j, \quad i = 1, \dots, N. \quad (5.5)$$

Determining the stability of the fixed point is then more involved mathematically than in the case of a single degree of freedom. We first note that Eq. (5.5) can be rewritten more formally in terms of the vector $\mathbf{E} = (\varepsilon_1, \dots, \varepsilon_N)^T$ (the superscript T denotes the matrix transpose) and the matrix \mathbf{M} of elements

$$M_{ij} = \frac{\partial F_i}{\partial x_j} \left(x_1^{(0)}, \dots, x_N^{(0)} \right). \quad (5.6)$$

With these notations, Eq. (5.5) then reads

$$\frac{d}{dt} \mathbf{E} = \mathbf{M} \mathbf{E}. \quad (5.7)$$

The matrix \mathbf{M} is sometimes called the stability matrix. Using the standard tools of linear algebra, the stability of the fixed point is given by the eigenvalue λ_M of the matrix \mathbf{M} having the largest real part, denoted as $\text{Re } \lambda_M$. If $\text{Re } \lambda_M < 0$, the fixed point is linearly stable, while if $\text{Re } \lambda_M > 0$ the fixed point is linearly unstable. The case $\text{Re } \lambda_M = 0$ corresponds to a marginally stable fixed point, whose actual stability is given by terms of higher order than the linear terms retained in Eq. (5.5). If the eigenvalues are all real, the linearized dynamics simply corresponds to a sum of exponential functions of time. If some of the eigenvalues are complex (they need to appear as pairs of complex conjugate values), then the linearized dynamics also includes oscillations.

A stable fixed point is actually the simplest example of the more general notion of attractor. An attractor is, generically speaking, a subset of phase space (i.e., the space in which the vector x_1, \dots, x_N is defined) onto which the dynamics concentrates at long time. These attractors can be classified according to their dimension. A stable fixed point is thus a zero-dimensional attractor, and a limit cycle is a one-dimensional attractor; attractors of higher dimension can also exist. Obviously, the dimension of the attractor cannot be larger than the dimension of phase space, that is the number N of dynamical variables. In some situations, the long time dynamics may be very irregular even at long times, and the dynamics is called chaotic in this case—we shall provide later on a more accurate definition of chaotic dynamics. For dissipative systems, chaotic dynamics leads to “strange attractors” having a fractal dimension. We shall come back to the description of chaotic dynamics in Sect. 5.1.3, though without describing strange attractors.

5.1.2 Bifurcations

When the dynamics of the system depends on an external control parameter, the stability of the fixed points, as well as their locations, generically depends on this control parameter. Most often, the location of the fixed point varies continuously with the control parameter, so that it is possible to “follow” the evolution of the fixed points. Moreover, a stable fixed point may become unstable when the control parameter is varied beyond a critical value. Such a change of stability, which is often accompanied by the onset of one or several new attractors (two stable fixed points, or a limit cycle for instance), is called a bifurcation. Let us illustrate this notion on

the simple example of the nonlinear dynamics of a single variable $x(t)$, described by an equation

$$\frac{dx}{dt} = f(x, \mu) \quad (5.8)$$

where μ is the control parameter. We assume for simplicity that $x = 0$ is a fixed point for all values of μ (otherwise, one simply needs to redefine x through a shift). This implies that $f(0, \mu) = 0$. We further assume that f is odd, meaning that $f(-x, \mu) = -f(x, \mu)$. Expanding $f(x, \mu)$ around $x = 0$, one generically obtains

$$f(x, \mu) = \alpha(\mu)x - \beta(\mu)x^3 + \mathcal{O}(x^5) \quad (5.9)$$

where $\alpha(\mu)$ and $\beta(\mu)$ are two functions of μ , that we assume for simplicity to be continuous. Note that even terms in x vanish due to the parity properties of f . A bifurcation occurs when there exists a value μ_0 such that $\alpha(\mu)$ changes sign at $\mu = \mu_0$. Without loss of generality, we assume that $\alpha(\mu) < 0$ for $\mu < \mu_0$ and $\alpha(\mu) > 0$ for $\mu > \mu_0$. Then, a linear stability analysis of the fixed point $x = 0$ leads to

$$\frac{dx}{dt} = \alpha(\mu)x. \quad (5.10)$$

Hence $x = 0$ is a stable fixed point for $\mu < \mu_0$, and an unstable fixed point for $\mu > \mu_0$, so that a bifurcation occurs for $\mu = \mu_0$. If the coefficient $\beta(\mu)$ appearing in Eq. (5.9) is strictly positive for $\mu \geq \mu_0$, a pair of symmetric fixed points appears when $\mu > \mu_0$ at $x = \pm x_0$, with

$$x_0 = \sqrt{\frac{\alpha(\mu)}{\beta(\mu)}} \quad (\alpha(\mu) > 0). \quad (5.11)$$

It can be checked easily that these new fixed points are linearly stable. The continuity of $\alpha(\mu)$ for $\mu \rightarrow \mu_0$ implies that the fixed points $\pm x_0$ continuously emerge from 0 when μ is increased above μ_0 (see Fig. 5.1 Left). Such a bifurcation is called a *supercritical bifurcation*.

If on the contrary $\beta(\mu) < 0$ for $\mu > \mu_0$, the determination of the emerging stable fixed points involves the term of order x^5 in the expansion given in Eq. (5.9), or more generally the lowest order stabilizing term. In this case, the new fixed points emerge at finite values $\pm x_0^*$ when μ exceeds a value $\mu^* < \mu_0$ (see Fig. 5.1 Right). Such a transition is called a *subcritical bifurcation*. Note that there are strong formal analogies between the simple bifurcations we have just presented, and the Landau theory of phase transitions, as described in Sect. 1.4. Note also that in systems with more than one degree of freedom, more complex bifurcations may occur, like the Hopf bifurcation in which a limit cycle appears when a fixed point becomes unstable as a control parameter is varied. Here again, notions of supercritical and subcritical bifurcations may be introduced depending on whether the limit cycle emerges continuously from a point, or with a finite size respectively.

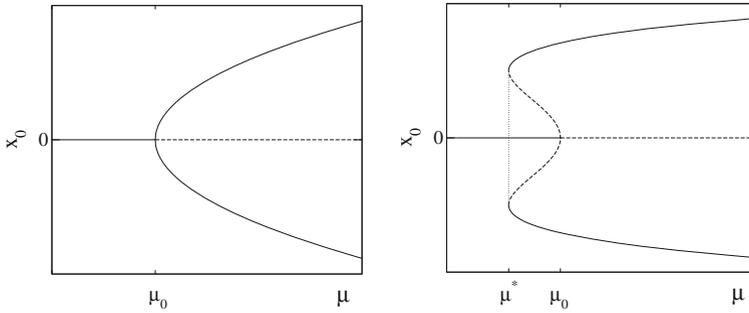


Fig. 5.1 Sketch of two standard types of bifurcations, occurring at a value μ_0 of the control parameter μ . Lines indicate the fixed points x_0 of the dynamics (*full line* stable fixed point; *dashed line* unstable fixed point). *Left* supercritical bifurcation; two symmetric stable fixed points appear continuously from $x_0 = 0$ for $\mu > \mu_0$. *Right* subcritical bifurcation; stable non-zero fixed points appear at a finite distance from $x_0 = 0$ for a value $\mu^* < \mu_0$, indicated by a vertical *dotted line*. Linear stability of the fixed point $x_0 = 0$ is lost only for $\mu > \mu_0$

5.1.3 Chaotic Dynamics

To discuss the notion of chaotic dynamics, it is actually more convenient to use the framework of discrete time dynamics. In this case, time takes only integer values and the value x_{t+1} of a dynamical variable at time $t + 1$ is given as a function of its value x_t at time t

$$x_{t+1} = f(x_t). \tag{5.12}$$

The function $f(x)$ is called a map, as it maps the interval of definition of the variable x onto itself. To emphasize the discreteness of time, we write it as a subindex. The discrete time dynamics may be interpreted as a periodic sampling of an underlying continuous time dynamics. Yet, this does not need to be the case, and a discrete time dynamics can also be considered in its own right, independently of any continuous time dynamics. The notion of fixed point, limit cycle and chaotic dynamics can be similarly defined for discrete time dynamics. A fixed point $x^{(0)}$ is such that

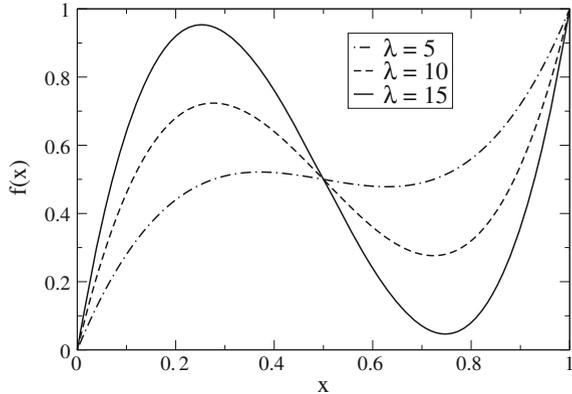
$$x^{(0)} = f(x^{(0)}). \tag{5.13}$$

The stability is tested by introducing a small perturbation around the fixed point, $x_t = x^{(0)} + \varepsilon_t$, yielding

$$\varepsilon_{t+1} = f'(x^{(0)}) \varepsilon_t. \tag{5.14}$$

The fixed point is thus linearly stable when ε converges to zero, that is when $|f'(x^{(0)})| < 1$. In the opposite case $|f'(x^{(0)})| > 1$, the fixed point is linearly unstable. In the simple case of the discrete time dynamics of a single degree of freedom as we consider here, a limit cycle consists in a finite set of q values (x_1^c, \dots, x_q^c) such that

Fig. 5.2 Illustration of the shape of the function $f(x)$ for $\lambda = 5$ (dot-dashed line), $\lambda = 10$ (dashed line) and $\lambda = 15$ (full line)



$$f(x_1^c) = x_2^c, \quad f(x_2^c) = x_3^c, \quad \dots, \quad f(x_q^c) = x_1^c. \quad (5.15)$$

If the dynamics does not converge in time to a fixed point or to a limit cycle, it may be chaotic, with an apparently erratic behavior. Chaoticity is defined by the fact that the distance between two initially close points increases exponentially with time; the growth rate is called the Lyapunov exponent. Note that this notion is different from the case of an unstable fixed point, since we are considering arbitrary closeby initial points, rather than the neighborhood of a fixed point. More generally, for a system with N degrees of freedom, there are N Lyapunov exponents. The system is chaotic if the largest Lyapunov exponent is positive.

As a simple illustration of the emergence of chaotic dynamics, let us consider the following map (see Fig. 5.2),

$$f(x) = \lambda x \left(x - \frac{1}{2} \right) (x - 1) + x \quad (5.16)$$

where λ is a parameter taken in the interval $0 < \lambda < 16$ to ensure that $0 < f(x) < 1$ for $0 < x < 1$ (negative values of λ would also satisfy this constraint in some range, but we focus here on positive values of λ). Fixed points, satisfying $f(x^{(0)}) = x^{(0)}$, are readily given by $x^{(0)} = 0, \frac{1}{2}$ and 1. The two fixed points $x^{(0)} = 0$ and $x^{(0)} = 1$ are unstable since $f'(0) = f'(1) = 1 + \frac{\lambda}{2} > 1$. The stability of the fixed point $x^{(0)} = \frac{1}{2}$ is more interesting as it depends on λ . We have

$$f' \left(\frac{1}{2} \right) = 1 - \frac{\lambda}{4} \quad (5.17)$$

so that $|f'(\frac{1}{2})| < 1$ for $0 < \lambda < 8$ and $|f'(\frac{1}{2})| > 1$ for $\lambda > 8$. As a result, the fixed point $x^{(0)} = \frac{1}{2}$ is linearly stable for $\lambda < 8$ and linearly unstable for $\lambda > 8$. In this latter case, we can check by numerical simulations towards which kind of attractor the dynamics converges. We see that for $\lambda = 10$, a limit cycle is obtained, while for

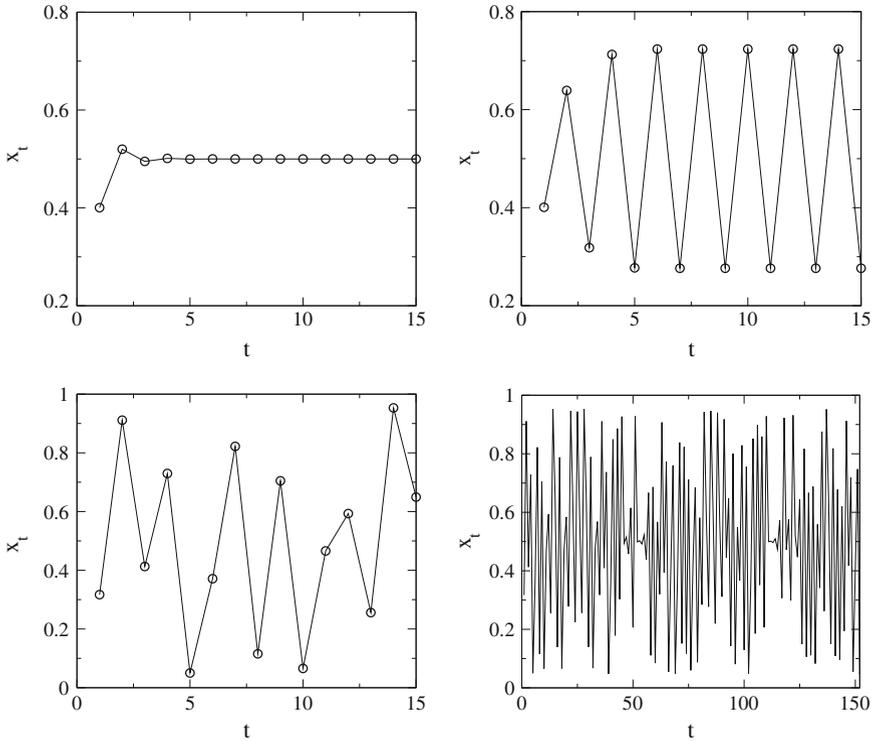


Fig. 5.3 *Top left* Convergence to a fixed point for $\lambda = 5$. *Top right* Convergence to a limit cycle (oscillation between two different points) for $\lambda = 10$. *Bottom* case $\lambda = 15$, showing a chaotic behaviour (*Left* same time window as on the top panels; *Right* larger time window)

$\lambda = 15$, a chaotic dynamics is observed—see Fig. 5.3. By chaotic, we simply mean here that the dynamics appears to be very irregular. A more quantitative statement would require a numerical evaluation of the Lyapunov exponent.

5.2 Deterministic Versus Stochastic Dynamics

5.2.1 Qualitative Differences and Similarities

An interesting notion to discuss qualitative similarities and differences between deterministic and stochastic dynamics is that of chaotic walk. Let us define a walk $y(t)$ through the relation

$$y_{t+1} = y_t + (2x_t - 1) \tag{5.18}$$

where the variable x_t , obeying Eq. (5.12), has been rescaled into $2x_t - 1$ so that it spans the entire interval $[-1, 1]$. An illustration of the chaotic walk is shown on the left panel of Fig. 5.4, and it turns out to be visually similar to a random walk, at least at first sight. Looking more carefully, one may however notice some anticorrelation between the steps, in the sense that positive steps are more often followed by negative steps than by positive ones. To make the comparison more quantitative, one can compute the mean displacement $\langle x_t \rangle$ and the mean square displacement $\langle x_t^2 \rangle$ —see right panel of Fig. 5.4. The average is taken over an ensemble of trajectories with different initial conditions, so that averages depend on time. More precisely, the initial position is given by $y_0 = 0$, and the initial value x_0 is uniformly sampled from the interval $(0, 1)$, taking a set of equidistant values over this interval. We observe that the mean displacement is almost equal to zero up to some small fluctuations, while the mean square displacement turns out to be linear in time, again up to some small fluctuations. These results coincide with the results obtained from a random walk, showing that random systems and chaotic systems share some common properties. In a sense, this is not very surprising since in practice, random processes are simulated on a computer using random number generators that are nothing but chaotic deterministic processes. However, let us stress that random number generators need to be tuned to satisfy required properties of statistical independence and uniformity of the generated numbers. We see here that taking an arbitrary map to compute a chaotic walk, we already obtain without any fine tuning some basic properties that are similar to that of random systems. In the following section, we discuss in more details this similarity between chaotic and random systems.

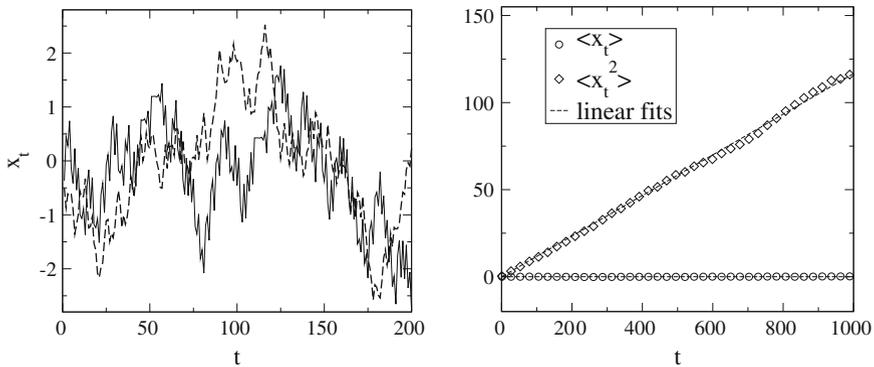


Fig. 5.4 *Left* Chaotic walk (*full line*); a random walk (*dashed line*) is shown for comparison. *Right* Mean displacement $\langle x_t \rangle$ and mean square displacement $\langle x_t^2 \rangle$ of the chaotic walk, obtained by averaging over many trajectories having different initial conditions. Similarly to what would be obtained for a random walk, the average displacement is zero for all time, and the mean square displacement increases linearly with time

5.2.2 *Stochastic Coarse-Grained Description of a Chaotic Map*

We now come back to the chaotic map $x_{t+1} = f(x_t)$ considered in Sect. 5.1.3. We have previously determined average values of a few observables in the chaotic regime of this map. Here we wish to go one step further and to determine the full histogram of the values of x_t . This histogram is displayed on Fig. 5.5 (left panel), for a value $\lambda = 15$ corresponding to a chaotic regime.

The histogram has been built following standard methods, namely dividing the interval $(0, 1)$ into a relatively large number of subintervals, called ‘bins’, and determining for a long trajectory the relative number of values of x_t that are contained in each bin. An important remark at this stage is that the dynamics of the map is purely deterministic only if the value of x_t is known with an infinite accuracy. To illustrate this issue, we may use the bins not only to build the histogram, but also to define an effective dynamics using coarse-grained configurations. In terms of these new configurations, the dynamics is no longer deterministic, because the evolution starting from a given bin at time t may lead to several distinct bins at time $t + 1$, while deterministic dynamics would require a single target bin. Stated otherwise, the knowledge of the initial coarse-grained configuration is not enough to determine the configuration at any later time. Interestingly, this property is also a characteristic property of stochastic systems, and we will now elaborate more quantitatively on this comparison.

To this aim, we define an auxiliary stochastic model as follows. Having defined a partition of the interval $(0, 1)$ into bins, we start by measuring the frequency of occurrence $F(j \rightarrow k)$ of direct (i.e., in a single step) transitions from bin j to bin k for any pair (j, k) . In a second stage, we define the auxiliary stochastic model as a Markov chain on the set of bins, with transition probabilities $T(j \rightarrow k)$ chosen to be equal to the frequencies $F(j \rightarrow k)$ measured in the chaotic dynamics. Simulating this Markov chain, we can also determine the corresponding empirical histogram of $x(t)$. The result is shown on the right panel of Fig. 5.5. A striking similarity is observed with the original histogram obtained from the chaotic dynamics. This similarity indicates that in practical situations, in which data necessarily have a finite resolution, it may be difficult to distinguish a deterministic process from a stochastic one.

These difficulties may actually be overcome by using more sophisticated tools to characterize the nature of the dynamics. For instance, under the assumption of a deterministic dynamics, one may try to characterize the dimension of the attractor. If this procedure is applied to a stochastic signal, the resulting dimension of the attractor would be found to be very large (in principle infinite). Yet, this type of analysis is, like the more naive ones previously discussed, limited in practice by the finite amount of available data (typically a finite set of points). To take this hard fact into account, an interesting proposition has been put forward, namely to characterize the deterministic or stochastic nature of a signal relatively to a given scale of resolution [1].

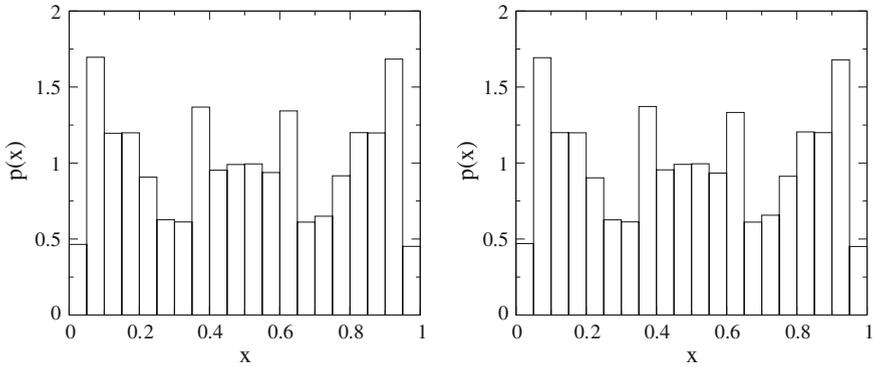


Fig. 5.5 *Left* Histogram of the values of x_t , using the deterministic evolution $x_{t+1} = f(x_t)$, in the case $\lambda = 15$. *Right* Histogram obtained from the effective stochastic process mimicking the deterministic one (see text). Both histograms turn out to be very similar

5.2.3 Statistical Description of Chaotic Systems

To go beyond the numerical analysis presented above, let us discuss the statistical approach to chaotic systems from a more theoretical perspective. Although chaotic systems are deterministic, they can be described by tools that are similar to that used for stochastic systems, namely probability distributions. In the case of deterministic systems, probabilistic aspects do not come from the evolution in itself, but rather from the fact that one follows an ensemble of trajectories, determined by a set of initial conditions. Hence one follows the evolution of a distribution $p_t(x)$ of configurations x as a function of time t , under the deterministic dynamics. We consider here the case of a deterministic dynamics defined by a map $x_{t+1} = f(x_t)$, and we assume without loss of generality that the variable x is defined over the interval $(0, 1)$ (any other interval, even unbounded, can be mapped onto the interval $(0, 1)$ through some—possibly nonlinear—transform). Formally, the evolution of the distribution $p_t(x)$ is given by the equation

$$p_{t+1}(x) = \int_0^1 dx' p_t(x') \delta(f(x') - x). \quad (5.19)$$

Note that the distribution $p_t(x)$ is normalized according to $\int_0^1 p(x) dx = 1$. Now one has to identify, for each value of x , the list of values x'_i , $i = 1, \dots, n(x)$ such that $f(x'_i) = x$. Then, thanks to the standard properties of Dirac delta functions (see Appendix A), one can rewrite

$$\delta(f(x') - x) = \sum_{i=1}^{n(x)} \frac{1}{|f'(x'_i)|} \delta(x' - x'_i) \quad (5.20)$$

so that Eq. (5.19) now reads, after integration of the Delta distributions,

$$p_{t+1}(x) = \sum_{i=1}^{n(x)} \frac{p_t(x'_i)}{|f'(x'_i)|}. \tag{5.21}$$

The stationary distribution is then obtained by assuming that $p_t(x)$ does not depend on t , yielding the following equation for the stationary distribution $p(x)$,

$$p(x) = \sum_{i=1}^{n(x)} \frac{p(x'_i)}{|f'(x'_i)|}. \tag{5.22}$$

This equation is in general complicated to solve, because it is non local (recall that x'_i is a function of x). In most cases, $n(x)$ is piecewise constant, which slightly simplifies the problem, although finding the solution on each interval where n is constant remains potentially difficult.

As a simple illustration, let us consider the following map, sometimes called (generalized) tent map:

$$f(x) = \begin{cases} \frac{x}{a} & \text{if } 0 \leq x \leq a, \\ \frac{1-x}{1-a} & \text{if } a \leq x \leq 1. \end{cases} \tag{5.23}$$

This map is illustrated in the left panel of Fig. 5.6. In this case, the equation $f(x'_i) = x$ has two solutions for all x (hence $n(x) = 2$). These solutions are given by

$$x'_1 = ax, \quad x'_2 = 1 - (1 - a)x \tag{5.24}$$

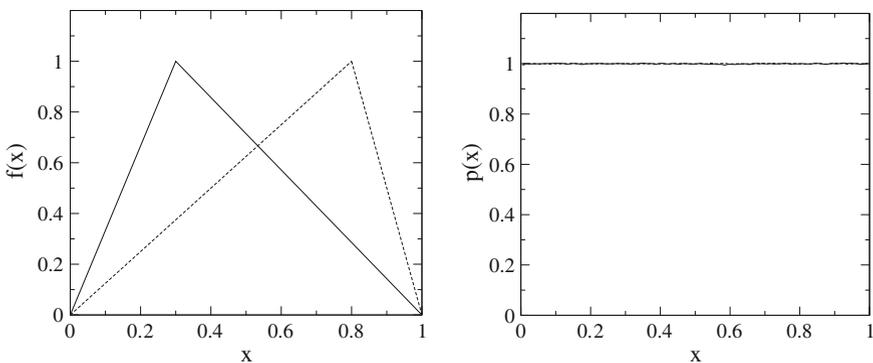


Fig. 5.6 *Left* Illustration of the shape of the tent map for $a = 0.3$ (full line) and $a = 0.8$ (dashed line). *Right* Stationary probability distribution measured numerically for the tent map given in Eq. (5.23), confirming the prediction of a uniform distribution (full line $a = 0.3$; dashed line $a = 0.8$). Both distributions are indistinguishable

and one has $|f'(x'_1)| = a$ and $|f'(x'_2)| = 1 - a$, so that Eq. (5.22) reads

$$p(x) = a p(ax) + (1 - a) p(1 - (1 - a)x). \quad (5.25)$$

A constant value of $p(x)$ is obviously a solution of this equation, and this constant has to be $p(x) = 1$ from the normalization condition $\int_0^1 p(x) dx = 1$. Note that the same uniform distribution is obtained whatever the value of a . Whether Eq. (5.25) admits other solutions is not easy to verify, but one can resort to numerical simulations to check whether this uniform distribution is the one asymptotically reached by the dynamics. Note that such a comparison requires the hypothesis of ergodicity, whereby ensemble averages are assumed to be equal to time averages. Under this assumption, we show on the right panel of Fig. 5.6 that the histogram computed from numerical simulations agrees with the predicted uniform distribution $p(x) = 1$, meaning either that the solution is unique, or that at least it is the dynamically selected solution.

The example has been chosen here to yield a simple enough analytic solution. Of course, and although the distribution is found here not to depend on the parameter a , not all maps yield a uniform stationary distribution. We have seen for instance in Fig. 5.5 that the map $f(x)$ given in Eq. (5.16) gives a nonuniform distribution.

5.3 Globally Coupled Dynamical Systems

5.3.1 Coupling Low-Dimensional Dynamical Systems

Up to now, we have discussed some basic properties of dynamical systems, focusing on low dimensional systems, that is, systems with a small number of degrees of freedom. An interesting generalization, much in the spirit of statistical physics, is to couple a large number of dynamical systems, each of them possibly having its own characteristics, as defined by some individual control parameter for instance. The coupled system is then a high dimensional dynamical system, that may exhibit a rich behavior. In the following, we restrict ourselves to large, globally coupled dynamical systems, in which a given individual dynamical system interacts with all the others in the same way. In principle, such a large system can also be studied purely in the framework of dynamical systems, by determining for instance the attractors in the high-dimensional phase space. However, such a task is in general very complicated. Also, in the spirit of statistical physics, one may wish to reduce the description of the system to a small number of global variables (or “order parameters”), based on the knowledge of the individual dynamical systems. A general procedure, involving some approximation scheme, has been developed to obtain such a reduced description in terms of a small number of average variables [2]. We will illustrate this procedure on the following globally coupled model, composed of N dynamical systems with two variables $x_j(t)$ and $y_j(t)$, for $j = 1, \dots, N$. The variables $x_j(t)$ and $y_j(t)$ obey the deterministic dynamics given by

$$\frac{dx_j}{dt} = \tau_j g(x_j, y_j) + k(X - x_j) \quad (5.26)$$

$$\frac{dy_j}{dt} = \tau_j h(x_j, y_j) + k(Y - y_j) \quad (5.27)$$

where $X(t) = N^{-1} \sum_{j=1}^N x_j(t)$ and $Y(t) = N^{-1} \sum_{j=1}^N y_j(t)$ are respectively the instantaneous average values of $x_j(t)$ and $y_j(t)$ over the population of dynamical systems. The functions $g(x, y)$ and $h(x, y)$ are at this stage arbitrary given functions, that do not depend on j . Heterogeneity is incorporated in the model by introducing a specific time scale $\tau_j > 0$ for each dynamical system j . The last terms in Eqs. (5.26) and (5.27) are global coupling terms, that constrain x_j and y_j to remain close to the population averages X and Y . The constant k is called the coupling constant.

In order to illustrate the usefulness of the order parameter expansion method, we will investigate how the coupling between dynamical systems may change the stability of the fixed point with respect to the uncoupled case $k = 0$ (or in other words, of the individual dynamics). To this aim, we assume that $(x, y) = (0, 0)$ is an unstable fixed point of the uncoupled system ($k = 0$). This implies in particular that $g(0, 0) = h(0, 0) = 0$, and that at least one of the two eigenvalues λ_1 and λ_2 of the stability matrix (see Sect. 5.1.1) has a positive real part. Then, for any value of the coupling k , the configuration $(x_j, y_j) = (0, 0)$, $j = 1, \dots, N$ (implying $X = Y = 0$) is also a fixed point of the dynamics. The question we wish to address now is whether this fixed point of the global system remains unstable when the coupling constant is increased.

5.3.2 Description in Terms of Global Order Parameters

A generic approximation method to obtain a closed set of equations for a reduced number of global order parameters is the following. One starts by expanding Eqs. (5.26) and (5.27) perturbatively into the deviations of the variables x_j , y_j and τ_j with respect to their population average values. Note that we assume that the coupling constant k is strong enough so that the global system is in a coherent regime where each individual system remains close to the population average value. We introduce the notations

$$x_j = X + \delta x_j, \quad y_j = Y + \delta y_j, \quad \tau_j = \bar{\tau} + \delta \tau_j \quad (5.28)$$

where $\bar{\tau} = N^{-1} \sum_{j=1}^N \tau_j$. Hence one has by definition $\sum_{j=1}^N \delta x_j = \sum_{j=1}^N \delta y_j = \sum_{j=1}^N \delta \tau_j = 0$. By appropriately rescaling the functions g and h , we can set $\bar{\tau} = 1$, without affecting the existence and the stability of the fixed point $(0, 0)$ of the individual dynamical systems. The order parameters consist in the population average parameters X and Y , and possibly a small number of other parameters that remain to be defined if needed. Expanding Eq. (5.26), one gets

$$\begin{aligned} \frac{dX}{dt} + \frac{d}{dt}\delta x_j &= g(X, Y) + g(X, Y) \delta\tau_j + \frac{\partial g}{\partial x}(X, Y) \delta x_j + \frac{\partial g}{\partial y}(X, Y) \delta y_j \\ &+ \frac{\partial g}{\partial x}(X, Y) \delta\tau_j \delta x_j + \frac{\partial g}{\partial y}(X, Y) \delta\tau_j \delta y_j - k\delta x_j \end{aligned} \quad (5.29)$$

where we have kept only linear terms with respect to δx_j and δy_j . A similar equation is obtained by expanding Eq. (5.27), simply exchanging δx_j and δy_j , and replacing g by h . Let us denote by $\overline{a_j}$ the population average of any quantity a_j , namely $\overline{a_j} \equiv N^{-1} \sum_{j=1}^N a_j$. Taking the population average of Eq. (5.29), we see that linear terms in $\delta\tau_j$, δx_j and δy_j cancel. Taking also into account the analogous equation for Y , we are thus left with the equations

$$\frac{dX}{dt} = g(X, Y) + \frac{\partial g}{\partial x}(X, Y) V + \frac{\partial g}{\partial y}(X, Y) W \quad (5.30)$$

$$\frac{dY}{dt} = h(X, Y) + \frac{\partial h}{\partial x}(X, Y) V + \frac{\partial h}{\partial y}(X, Y) W \quad (5.31)$$

where we have introduced the new order parameters

$$V = \overline{\delta\tau_j \delta x_j}, \quad W = \overline{\delta\tau_j \delta y_j}. \quad (5.32)$$

To close the set of equations (5.30) and (5.31), one needs to find evolution equations for the new global variables V and W . The equation for V is obtained by multiplying Eq. (5.29) by $\delta\tau_j$ and averaging over the population. Applying a similar procedure to obtain an equation for W , we eventually get

$$\frac{dV}{dt} = \sigma^2 g(X, Y) + \frac{\partial g}{\partial x}(X, Y) V + \frac{\partial g}{\partial y}(X, Y) W - kV \quad (5.33)$$

$$\frac{dW}{dt} = \sigma^2 h(X, Y) + \frac{\partial h}{\partial x}(X, Y) V + \frac{\partial h}{\partial y}(X, Y) W - kW \quad (5.34)$$

where $\sigma^2 = \overline{\delta\tau_j^2}$. Hence we have obtained a closed set of four equations for the four order parameters X , Y , V and W . Note that to close the equations, we have neglected terms proportional to $\overline{\delta\tau_j^2 \delta x_j}$ or $\overline{\delta\tau_j^2 \delta y_j}$.

5.3.3 Stability of the Fixed Point of the Global System

Having obtained Eqs. (5.30), (5.31), (5.33) and (5.34), we can now proceed to the stability analysis of the global fixed point $(X, Y, V, W) = (0, 0, 0, 0)$, corresponding to a situation in which all individual dynamical systems are at the point $(x_j, y_j) = (0, 0)$. We thus linearize the set of Eqs. (5.30), (5.31), (5.33) and (5.34), yielding a matrix equation

$$\frac{d}{dt}\mathbf{Z} = \mathbf{M}\mathbf{Z} \quad (5.35)$$

where we have introduced the vector $\mathbf{Z} = (X, Y, V, W)^T$, and \mathbf{M} is the stability matrix. To lighten notations, we define

$$g_1 = \frac{\partial g}{\partial x}(0, 0), \quad g_2 = \frac{\partial g}{\partial y}(0, 0), \quad h_1 = \frac{\partial h}{\partial x}(0, 0), \quad h_2 = \frac{\partial h}{\partial y}(0, 0). \quad (5.36)$$

The stability matrix \mathbf{M} is then given by

$$\mathbf{M} = \begin{pmatrix} g_1 & g_2 & g_1 & g_2 \\ h_1 & h_2 & h_1 & h_2 \\ \sigma^2 g_1 & \sigma^2 g_2 & g_1 - k & g_2 \\ \sigma^2 h_1 & \sigma^2 h_2 & h_1 & h_2 - k \end{pmatrix}. \quad (5.37)$$

One needs to compute the eigenvalues of the matrix \mathbf{M} in order to determine the stability of the global system. Interestingly, one sees that \mathbf{M} has a natural block structure in terms of the 2×2 matrix \mathbf{M}_2 which describes the linear stability of the individual, uncoupled dynamical system (x, y) , namely

$$\mathbf{M}_2 = \begin{pmatrix} g_1 & g_2 \\ h_1 & h_2 \end{pmatrix}. \quad (5.38)$$

Using this block matrix, \mathbf{M} reads

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_2 & \mathbf{M}_2 \\ \sigma^2 \mathbf{M}_2 & \mathbf{M}_2 - k \mathbf{I}_2 \end{pmatrix}, \quad (5.39)$$

with \mathbf{I}_2 the two-dimensional identity matrix. Diagonalizing \mathbf{M}_2 , there exists a matrix \mathbf{P}_2 such that

$$\mathbf{P}_2^{-1} \mathbf{M}_2 \mathbf{P}_2 = \mathbf{D}_2 \equiv \begin{pmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{pmatrix} \quad (5.40)$$

where the eigenvalues ν_1 and ν_2 are the two solutions of the equation

$$v^2 - (g_1 + h_2)v + g_1 h_2 - h_1 g_2 = 0. \quad (5.41)$$

Note that in what follows, we actually do not need to determine the matrix \mathbf{P}_2 explicitly. Let us introduce the 4×4 matrix \mathbf{Q} defined in terms of blocks as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{P}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 \end{pmatrix}. \quad (5.42)$$

Its inverse matrix \mathbf{Q}^{-1} is simply given by

$$\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{P}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2^{-1} \end{pmatrix}. \quad (5.43)$$

Now we can compute the matrix $\tilde{\mathbf{M}} = \mathbf{Q}^{-1}\mathbf{M}\mathbf{Q}$, which reads

$$\tilde{\mathbf{M}} = \begin{pmatrix} \mathbf{D}_2 & \mathbf{D}_2 \\ \sigma^2\mathbf{D}_2 & \mathbf{D}_2 - k\mathbf{I}_2 \end{pmatrix}, \quad (5.44)$$

or more explicitly

$$\tilde{\mathbf{M}} = \begin{pmatrix} v_1 & 0 & v_1 & 0 \\ 0 & v_2 & 0 & v_2 \\ \sigma^2 v_1 & 0 & v_1 - k & 0 \\ 0 & \sigma^2 v_2 & 0 & v_2 - k \end{pmatrix}. \quad (5.45)$$

Since $\tilde{\mathbf{M}}$ and \mathbf{M} are related by a similarity transform, they share the same eigenvalues. The eigenvalues λ of $\tilde{\mathbf{M}}$ are determined by solving the equation $\det(\tilde{\mathbf{M}} - \lambda\mathbf{I}) = 0$, with \mathbf{I} the 4×4 identity matrix. After some straightforward algebra, one finds

$$\det(\tilde{\mathbf{M}} - \lambda\mathbf{I}) = [(v_1 - \lambda)(v_1 - k - \lambda) - \sigma^2 v_1][(v_2 - \lambda)(v_2 - k - \lambda) - \sigma^2 v_2]. \quad (5.46)$$

Hence the eigenvalues are solutions of one of the following equations, for $i = 1$ or 2 ,

$$(v_i - \lambda)(v_i - k - \lambda) - \sigma^2 v_i = 0, \quad (5.47)$$

yielding the four eigenvalues

$$\lambda_{i,\pm} = v_i - \frac{k}{2} \pm \sqrt{\frac{k^2}{4} + \sigma^2 v_i^2}, \quad i = 1, 2. \quad (5.48)$$

Note that the quantity under the square root may be a complex number. We are interested in checking whether the presence of the coupling between individual dynamical systems may restabilize the fixed point $X = Y = 0$, which is assumed to be unstable in individual systems. The stability criterion for the global system is that the four eigenvalues $\lambda_{i,\pm}$ have a negative real part. It turns out that this happens when the eigenvalues v_i are complex (and thus necessarily complex conjugate) [2], so that we set $v_{1,2} = \beta \pm i\gamma$ (β, γ real), with $\beta > 0$ due to the assumed instability of the fixed point for individual uncoupled systems. The stability criterion $\text{Re } \lambda_{i,\pm} < 0$ can be reexpressed as

$$\left| \text{Re} \sqrt{\frac{k^2}{4} + \sigma^2(\beta + i\gamma)^2} \right| < \frac{k}{2} - \beta. \quad (5.49)$$

One can immediately check that if $\gamma = 0$, condition (5.49) cannot be fulfilled, so that complex eigenvalues $\nu_{1,2}$ are indeed required. It also appears clearly that heterogeneity is needed too: if all the dynamical systems share the same time constant τ_j , one has $\sigma = 0$ and condition (5.49) cannot hold. Assuming $\gamma \gg \beta$, and performing asymptotic expansions of Eq. (5.49) for both $k \ll \gamma$ and $k \gg \gamma$, one finds that the fixed point $(X, Y, V, W) = (0, 0, 0, 0)$ is stable in the range of coupling values

$$2\beta(1 + \sigma) < k < (\gamma^2 - \beta^2) \frac{\sigma^2}{\beta}. \quad (5.50)$$

Hence a sufficiently strong coupling (together with the presence of heterogeneity) is able to restabilize the unstable fixed point present in individual uncoupled systems. However, it turns out that for very large values of the coupling, the global fixed point is again unstable. This restabilization effect has been observed in different types of dynamical systems where the dynamics converges to a limit cycle after getting away from the unstable fixed point. The phenomenon of restabilization of the fixed point when increasing the coupling has been called “oscillator death” in this context [2].

5.4 Synchronization Transition

We now turn to another phenomenon emerging from the global coupling of low dimensional deterministic systems, namely the synchronization transition, through which coupled oscillators with distinct natural frequencies oscillate in phase, with a common frequency, if the coupling is strong enough. The paradigmatic model for the synchronization transition is the Kuramoto model, that we describe below.

5.4.1 The Kuramoto Model of Coupled Oscillators

The Kuramoto model [3] consists in a set of N oscillators of phase θ_j , evolving according to the coupled equations

$$\frac{d\theta_j}{dt} = \omega_j + \sum_{k=1}^N K_{jk} \sin(\theta_k - \theta_j), \quad j = 1, \dots, N, \quad (5.51)$$

where ω_j is the natural frequency of oscillator j , and K_{jk} is the coupling constant between oscillators j and k . Applications of the Kuramoto model range from chemical oscillators to neural networks, laser arrays or Josephson junctions [4]. We shall here mostly follow the presentation of this model given in Ref. [4], and refer the reader to this specialized review for further details.

The most simple version of the Kuramoto model is obtained by choosing uniform (mean-field type) couplings $K_{jk} = K/N$, such that any pair of oscillators has the same coupling. The $1/N$ scaling is included so that the sum of all coupling terms does not trivially dominate the natural frequency in Eq. (5.51). The evolution of θ_j is then given by

$$\frac{d\theta_j}{dt} = \omega_j + \frac{K}{N} \sum_{k=1}^N \sin(\theta_k - \theta_j), \quad j = 1, \dots, N. \quad (5.52)$$

In order to characterize the possible synchronization of the oscillators resulting from the coupling terms, it is convenient to introduce the complex order parameter $r e^{i\psi}$ defined as

$$r e^{i\psi} = \frac{1}{N} \sum_{k=1}^N e^{i\theta_k}. \quad (5.53)$$

In the absence of synchronization, the (mean) value of this order parameter is equal to zero, while the presence of synchronization is indicated by a value $r > 0$, the phase ψ corresponding to the ‘average’ phase of the oscillators. It is convenient to reformulate Eq. (5.52) as

$$\frac{d\theta_j}{dt} = \omega_j + Kr \sin(\psi - \theta_j), \quad j = 1, \dots, N, \quad (5.54)$$

using the fact that from Eq. (5.53),

$$r e^{i(\psi - \theta_j)} = \frac{1}{N} \sum_{k=1}^N e^{i(\theta_k - \theta_j)} \quad (5.55)$$

for any j , and taking the imaginary part of Eq. (5.55).

We shall now focus on the limit of a very large number of coupled oscillators, $N \rightarrow \infty$. In this case, the natural frequencies are described by the density $g(\omega)$, which means that the fraction of oscillators having a natural frequency ω_j in the infinitesimal range $[\omega, \omega + d\omega]$ is $g(\omega)d\omega$. The density $g(\omega)$ is normalized as $\int_{-\infty}^{\infty} g(\omega)d\omega = 1$. By an appropriate transform $\theta \mapsto \theta - \Omega t$, it is possible to redefine the model in such a way that $\langle \omega \rangle \equiv \int_{-\infty}^{\infty} \omega g(\omega)d\omega = 0$. This can be interpreted as looking at the oscillators in a rotating frame. In this frame, synchronization appears as a time independent value of the average phase ψ , with $r > 0$.

The statistics of the phases of oscillators having a given frequency ω is encoded into the time-dependent probability distribution $\rho(\theta|\omega, t)$. This distribution, normalized according to $\int_{-\infty}^{\infty} \rho(\theta|\omega, t)d\theta = 1$, describes the statistics of a set of identical oscillators having different initial conditions. Taking into account Eq. (5.54), the evolution of the distribution $\rho(\theta|\omega, t)$ is governed by the equation¹

¹This equation may be thought of as a Fokker-Planck equation (see Sect. 2.3) in the zero noise limit.

$$\frac{\partial \rho}{\partial t}(\theta|\omega, t) + \frac{\partial}{\partial \theta} \left[\left(\omega + Kr \sin(\psi - \theta) \right) \rho(\theta|\omega, t) \right] = 0. \quad (5.56)$$

In the infinite N limit considered here, the expression (5.53) of the order parameter reduces to

$$r e^{i\psi} = \langle e^{i\theta} \rangle \equiv \int_{-\pi}^{\pi} d\theta \int_{-\infty}^{\infty} d\omega e^{i\theta} \rho(\theta|\omega, t) g(\omega). \quad (5.57)$$

In the following, we look for steady-state solutions and study whether the oscillators get synchronized or not in this regime, depending on the coupling strength K .

5.4.2 Synchronized Steady State

In order to find the steady-state solution of the model, we need to find for all frequency ω the time-independent distribution $\rho(\theta|\omega)$ solution of Eq. (5.56), in which r and ψ are time-independent values self-consistently determined from Eq. (5.57). It can easily be checked that the uniform distribution $\rho(\theta|\omega) = (2\pi)^{-1}$, which leads to $r = 0$, is a solution of Eq. (5.56) for all coupling strength K . This solution corresponds to a complete lack of synchronization between oscillators. While such a situation is likely to be relevant at low coupling, it is however possible that other solutions exist if the coupling strength K is strong enough.

To look for such possible solutions, we start from a given value of the order parameter $r e^{i\psi}$ with $r > 0$, determine the solution of Eq. (5.56) for these values of r and ψ , and then check whether a self-consistent solution of Eq. (5.57) can be found. We first note that if a stationary solution with global phase ψ exists, then another steady-state solution of phase $\psi + \alpha$ can be obtained by shifting all the phases θ_j by the same amount α . Hence we can restrict our study to the case $\psi = 0$, the other cases being deduced by a simple phase shift.

Under this assumption, the steady-state solution of Eq. (5.56) satisfies

$$\left(\omega - Kr \sin \theta \right) \rho(\theta|\omega) = C \quad (5.58)$$

where C is a constant. The condition $\rho(\theta|\omega) \geq 0$ implies that such a solution exists only if $|\omega| \geq Kr$. The case $|\omega| = Kr$ is further excluded as it would lead to a non-normalizable distribution $\rho(\theta|\omega)$. As a result, one finds

$$\rho(\theta|\omega) = \frac{1}{2\pi} \frac{\sqrt{\omega^2 - (Kr)^2}}{|\omega - Kr \sin \theta|}, \quad |\omega| > Kr. \quad (5.59)$$

If $|\omega| \leq Kr$, the distribution (5.59) is no longer valid. We leave aside the discussion of the marginal case $|\omega| = Kr$, which plays no role in the following, and focus on the situation $|\omega| < Kr$. In this case, the evolution equation (5.54) has two fixed points, solutions of

$$\omega - Kr \sin \theta = 0. \quad (5.60)$$

To check the linear stability of a fixed point θ_0 , we set $\theta = \theta_0 + \epsilon$, with $\epsilon \ll 1$. Expanding Eq. (5.54) to linear order in ϵ , we get

$$\frac{d\epsilon}{dt} = -(Kr \cos \theta_0) \epsilon, \quad (5.61)$$

so that the fixed point θ_0 is stable if $\cos \theta_0 > 0$ and unstable if $\cos \theta_0 < 0$. Taking into account Eq. (5.60), the stable fixed point is thus given by

$$\theta_0 = \sin^{-1} \left(\frac{\omega}{Kr} \right). \quad (5.62)$$

The distribution $\rho(\theta|\omega)$ associated to this fixed point solution is a Dirac delta function (see Appendix A), that is an infinitely peaked solution around the fixed point:

$$\rho(\theta|\omega) = \delta \left(\theta - \sin^{-1}(\omega/Kr) \right), \quad |\omega| < Kr. \quad (5.63)$$

Now that we have determined $\rho(\theta|\omega)$ for both $|\omega| < Kr$ and $|\omega| > Kr$, we can self-consistently determine r from Eq. (5.57), setting $\psi = 0$:

$$\begin{aligned} r = & \int_{-\pi}^{\pi} d\theta \int_{-Kr}^{Kr} d\omega e^{i\theta} \delta \left(\theta - \sin^{-1}(\omega/Kr) \right) g(\omega) \\ & + \frac{\sqrt{\omega^2 - (Kr)^2}}{2\pi} \int_{-\pi}^{\pi} d\theta \int_{|\omega|>Kr} d\omega \frac{e^{i\theta} g(\omega)}{|\omega - Kr \sin \theta|}. \end{aligned} \quad (5.64)$$

Let us now further assume that $g(\omega)$ is an even function, that is $g(-\omega) = g(\omega)$ for all ω (which is consistent with the assumption $\langle \omega \rangle = 0$). Using the symmetries of the sine function, it can be shown that the second integral in Eq. (5.64) is equal to zero. The first integral can be computed thanks to the properties of the δ function, namely

$$\int_a^b dx f(x) \delta(x - x_0) = f(x_0) \quad (5.65)$$

for any function f , provided that $a < x_0 < b$. One thus finds, exchanging the order of integration between θ and ω :

$$r = \int_{-Kr}^{Kr} d\omega g(\omega) e^{i \sin^{-1}(\omega/Kr)}. \quad (5.66)$$

Using the parity of $g(\omega)$, the imaginary part of the integral vanishes, and Eq. (5.66) reduces to

$$r = 2 \int_0^{Kr} d\omega g(\omega) \cos \left(\sin^{-1}(\omega/Kr) \right). \quad (5.67)$$

Performing the change of variable $\omega = Kr \sin x$, one eventually finds the following self-consistent equation, taking into account the assumption $r > 0$

$$\int_0^{\pi/2} dx (\cos x)^2 g(Kr \sin x) = \frac{1}{2K}. \quad (5.68)$$

The solutions of this equation depend on some generic properties of the function $g(\omega)$. In the following, we assume that $g(\omega)$ has its maximum at $\omega = 0$, that is for all $\omega \neq 0$, $g(\omega) < g(0)$. Denoting as $I(r)$ the integral on the left-hand-side of Eq. (5.68), we have for all $r > 0$, $I(r) < I(0)$. Hence if the coupling constant K is such that $(2K)^{-1} > I(0)$, Eq. (5.68) has no solution for r , while a solution $r > 0$ exists for $(2K)^{-1} < I(0)$. This defines the critical coupling $K_c = [2I(0)]^{-1}$, above which a solution $r > 0$ exists. Expanding $g(\omega)$ for small ω as

$$g(\omega) = g(0) - \frac{1}{2}|g''(0)|\omega^2 + \mathcal{O}(\omega^4), \quad (5.69)$$

with $g''(0) < 0$, one finds after some algebra the following relation, for $0 < K - K_c \ll K_c$,

$$r \approx \sqrt{\frac{16(K - K_c)}{\pi K_c^4 |g''(0)|}}. \quad (5.70)$$

The above result is valid for any regular function $g(\omega)$ having its maximum at $\omega = 0$. In the specific case

$$g(\omega) = \frac{1}{\pi} \frac{\omega_0}{\omega_0^2 + \omega^2}, \quad (5.71)$$

where $\omega_0 > 0$ is a constant, the solution of Eq. (5.68) can be given explicitly for all r , namely:

$$r = \sqrt{1 - \frac{2\omega_0}{K}}. \quad (5.72)$$

Finally, it can be shown that the synchronized solution $r > 0$ corresponds to the stable state of the system for $K > K_c$ [4].

References

1. M. Cencini, M. Falcioni, E. Olbrich, H. Kantz, A. Vulpiani, Chaos or noise: difficulties of a distinction. *Phys. Rev. E* **62**, 427 (2000)
2. S. De Monte, F. d'Ovidio, E. Mosekilde, Coherent regimes of globally coupled dynamical systems. *Phys. Rev. Lett.* **90**, 054102 (2003)
3. Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence* (Springer, New York, 1984)
4. J.A. Acebrón, L.L. Bonilla, C.J. Pérez, Vicente, F. Ritort, R. Spigler, The Kuramoto model: a simple paradigm for synchronization phenomena. *Rev. Modern Phys.* **77**, 137 (2005)

Chapter 6

A Probabilistic Viewpoint on Fluctuations and Rare Events

Statistical physics obviously bears some strong connection with probability theory. Indeed, the very basis of statistical physics is to associate a probability to each microscopic configuration of a system. When dealing with dynamics, the mathematical theory of stochastic Markov processes also plays a central role, as we have seen in Sect. 2.1. In this chapter, we wish to introduce some further aspects of probability theory which are of relevance to statistical physics, namely the properties of the sum and of the extreme values of a set of random variables.

6.1 Global Fluctuations as a Random Sum Problem

A generic problem of interest in statistical physics is to determine the statistics of fluctuating global observables, like the total energy, magnetization or number of particles, which are obtained as the sum of a large number of individual contributions associated to small regions of the system. Probabilistic theorems describing the behavior of random sums are thus of great importance in statistical physics.

6.1.1 *Law of Large Numbers and Central Limit Theorem*

We start by discussing two cornerstones of the probabilistic theory of random sums, namely the Law of Large Numbers and the Central Limit Theorem. Both of them deal with the statistical properties of sums of random variables. Roughly speaking, the Law of Large Numbers describes the fact that the empirical average of a series of random variables converges to the theoretical average (the expectation, in probabilistic terms). On the other side, the Central Limit Theorem characterizes the tiny fluctuations of the empirical average around the expectation.

To formulate these two theorems, we consider a set of independent and identically distributed random variables (x_1, \dots, x_N) , drawn from a common distribution $p(x)$, and we define the sum $S_N = \sum_{j=1}^N x_j$.

Law of Large Numbers. Let us define the rescaled sum $s_N = S_N/N$, which is nothing but the empirical average of the set of variables (x_1, \dots, x_N) . The probability distribution of s is denoted as $\Phi_N(s)$. Under the assumption that the distribution $p(x)$ has a finite average value $m = \int x p(x) dx$, the Law of Large Numbers states that the distribution $\Phi_N(s)$ converges to a Dirac distribution,

$$\Phi_N(s) \rightarrow \delta(s - m), \quad N \rightarrow \infty \quad (6.1)$$

(see Ref. [1] for a mathematically more rigorous formulation). In other words, the random variable s_N converges to the non-random value m when $N \rightarrow \infty$.

The Law of Large Numbers has important applications in statistical physics. The fact that global observables like the total energy or total magnetization in a large system composed of many degrees of freedom has only tiny fluctuations precisely comes from the Law of Large Numbers. In addition, the fact that the probability of an event can be measured as the empirical frequency of appearance of the event over a large number of independent realizations is also a consequence of the Law of Large Numbers. As an example, let us briefly discuss how the cumulative distribution function $F(x_0) = \int_{-\infty}^{x_0} p(x) dx$ of a random variable x can be evaluated in this way. One starts by introducing an auxiliary variable $y = \theta(x_0 - x)$, where x_0 is an arbitrary constant, and θ is the Heaviside function, equal to 1 for a positive argument and to 0 otherwise. Applying the Law of Large Numbers to the variable y , one finds that the empirical average $s_N = N^{-1} \sum_{j=1}^N \theta(x_0 - x_j)$ converges to $\langle \theta(x_0 - x) \rangle = F(x_0)$ when $N \rightarrow \infty$. Since this result is valid for any x_0 , s_N is thus an estimator of the cumulative probability distribution $F(x_0)$.

Central Limit Theorem. Knowing that the empirical average s_N converges to its expectation m , one can wonder about the amplitude and the shape of the fluctuations of s_N around m . Assuming that the distribution $p(x)$ has a finite second moment $\langle x^2 \rangle = \int x^2 p(x) dx$, one easily finds that the variance of s_N is equal to σ^2/N , where σ^2 is the variance of x (note that a finite second moment also implies a finite first moment). Fluctuations of s_N around m are thus of the order of $1/\sqrt{N}$.

To characterize the shape of these fluctuations (or equivalently, of the fluctuations of $S_N = \sum_{j=1}^N x_j$ around its expectation Nm), we introduce the rescaled variable

$$z_N = \frac{s_N - m}{1/\sqrt{N}} = \frac{S_N - Nm}{\sqrt{N}}, \quad (6.2)$$

which by definition has a variance equal to σ^2 . The Central Limit Theorem states that the distribution $\Psi_N(z)$ of the variable z_N converges to a centered Gaussian distribution of variance σ^2 ,

$$\Psi_N(z) \rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}, \quad N \rightarrow \infty. \quad (6.3)$$

Let us emphasize that both the Law of Large Numbers and the Central Limit Theorem rely on the assumptions that the variables are independent, identically distributed, and have a finite first moment, as well as a finite second moment in the case of the Central Limit Theorem. When at least one of these assumptions breaks down, the theorems are no longer valid. In practice, the convergence results given in Eqs. (6.1) and (6.3) remain valid when the random variables x_i 's are not too strongly correlated, and have distributions that do not strongly differ one from the other. In the case of strongly correlated variables, and/or strongly non-identically distributed random variables, the limit distributions may differ from that given in Eqs. (6.1) and (6.3). A scaling different from Eq. (6.2) may also be needed in some cases to obtain a convergence to a well-defined limit distribution. This is also true when the assumption of finite variance is broken: as soon as σ^2 is infinite, the limit distribution becomes non-Gaussian, and a scaling different from Eq. (6.2) is needed to reach a limit distribution. This situation is described by the Generalized Central Limit Theorem, that we now present.

6.1.2 Generalization to Variables with Infinite Variances

As we have just mentioned, a generalization of the Central Limit Theorem is required when the variables considered have an infinite variance. Before stating the Generalized Central Limit Theorem, let us first provide typical examples of probability distributions having infinite variances. Such laws typically have (at least approximately) a power-law tail, and for the sake of simplicity, we briefly discuss here only the case of a pure power-law distribution

$$p(x) = \frac{\alpha x_0^\alpha}{x^{1+\alpha}}, \quad x \geq x_0 \quad (\alpha > 0) \quad (6.4)$$

with $p(x) = 0$ for $x < x_0$ (a lower bound is necessary to make the distribution normalizable). A distribution like Eq. (6.4) is sometimes called a Pareto distribution. The second moment of the distribution reads

$$\langle x^2 \rangle = \int_{x_0}^{\infty} x^2 p(x) dx = \int_{x_0}^{\infty} \frac{\alpha x_0^\alpha}{x^{\alpha-1}} dx \quad (6.5)$$

which converges on condition that $\alpha > 2$. If $\alpha \leq 2$, the second moment $\langle x^2 \rangle$ is infinite. As a result, the Central Limit Theorem only applies to sets of independent and identically distributed random variables (x_1, \dots, x_N) distributed according to Eq. (6.4) if $\alpha > 2$. For $\alpha \leq 2$ a generalization of the theorem is needed.

Before presenting the Generalized Central Limit Theorem, we first need to introduce the Lévy distribution $L(z; \alpha, \beta)$, which is defined for $0 < \alpha \leq 2$ and $-1 \leq \beta \leq 1$ through its characteristic function (that is, the Fourier transform of the probability density)

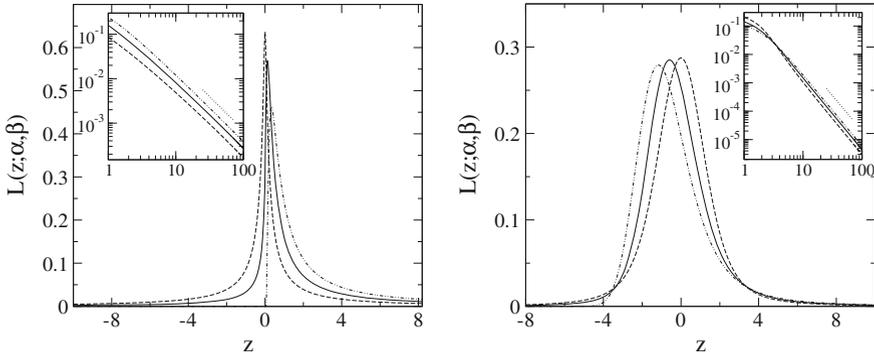


Fig. 6.1 Illustration of the Lévy distribution $L(z; \alpha, \beta)$ for different values of α and β . *Left* $\alpha = 0.5$, and $\beta = 0$ (*dashed line*), 0.5 (*full line*) and 1 (*dot-dashed*). The *inset* shows the positive tails of these distributions, which behave as $1/x^{1+\alpha}$ (the *dotted line* indicates a slope -1.5). *Right* $\alpha = 1.5$, and $\beta = 0$ (*dashed line*), 0.5 (*full line*) and 1 (*dot-dashed*). *Inset* positive tails of the same distributions (*dotted line* slope -2.5)

$$\begin{aligned} \hat{L}(k; \alpha, \beta) &\equiv \int_{-\infty}^{\infty} L(z; \alpha, \beta) e^{ikz} dz \\ &= \exp \left[-|k|^\alpha \left(1 - i\beta \operatorname{sgn}(k) \varphi(k, \alpha) \right) \right] \end{aligned} \quad (6.6)$$

with

$$\varphi(k, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2} & \text{if } \alpha \neq 1, \\ \frac{2}{\pi} \ln |k| & \text{if } \alpha = 1. \end{cases} \quad (6.7)$$

The Lévy distribution $L(z; \alpha, \beta)$ is illustrated in Fig. 6.1, for $\beta \geq 0$. Distributions with $\beta < 0$ are the symmetric of the ones for $\beta > 0$ with respect to the Y -axis, since $L(z; \alpha, -\beta) = L(-z; \alpha, \beta)$. In practice, the Lévy distribution $L(z; \alpha, \beta)$ is obtained by inverting the Fourier transform (6.6), leading to

$$L(z; \alpha, \beta) = \frac{1}{\pi} \int_0^{\infty} dk e^{-k^\alpha} \cos(\beta k^\alpha \varphi(k, \alpha) - kz). \quad (6.8)$$

The integral in Eq. (6.8) then has to be evaluated numerically.

We now provide the formulation of the Generalized Central Limit Theorem. We denote again as (x_1, \dots, x_N) a set of N independent and identically distributed random variables drawn from a distribution $p(x)$ having an infinite second moment $\int_{-\infty}^{\infty} x^2 p(x) dx$. The cumulative distribution function is defined as $F(x) \equiv \int_{-\infty}^x p(x') dx'$. We define for a given set of constants $\{a_N\}$ and $\{b_N\}$ the rescaled sum of the variables x_i ,

$$z_N = \frac{1}{b_N} \left(\sum_{j=1}^N x_j - a_N \right). \quad (6.9)$$

The Generalized Central Limit Theorem [1, 2] states that for a suitable choice of the rescaling parameters a_N and b_N , the distribution $\Psi_N(z)$ of the rescaled sum z_N converges to the Lévy distribution $L(z; \alpha, \beta)$ if the following conditions are satisfied

$$\lim_{x \rightarrow \infty} \frac{F(-x)}{1 - F(x)} = \frac{1 - \beta}{1 + \beta} \quad (6.10)$$

$$\forall r > 0, \quad \lim_{x \rightarrow \infty} \frac{1 - F(x) + F(-x)}{1 - F(rx) + F(-rx)} = r^\alpha. \quad (6.11)$$

Let us now briefly comment on this Generalized Central Limit Theorem. First, it is worth noticing that the parameter α in the Lévy distribution $L(z; \alpha, \beta)$ has the same interpretation as in the example of the Pareto distribution discussed in Eq. (6.4): it characterizes the power-law decay of the tail of the distribution. In fact, both the Lévy distribution and the distribution $p(x)$ from which the summed variables are drawn decay at large x as a power law with exponent $1 + \alpha$. The parameter β , on the other side, characterizes the asymmetry of the Lévy distribution. A value $\beta = 0$ corresponds to a symmetric distribution $L(-z; \alpha, 0) = L(z; \alpha, 0)$, while for $\beta > 0$ (resp. $\beta < 0$) the positive (resp. negative) tail carries a higher probability weight than the opposite tail.

The parameters α and β of the limit distribution can be determined from the original distribution $p(x)$, according to Eqs. (6.10) and (6.11). In practice, simpler criteria can however be used—see below. Let us first notice that the distribution $p(x)$ actually has two tails, one in $+\infty$ and one in $-\infty$. The parameter α is related to the tail with the slowest decay. Let us assume that $p(x)$ behaves as

$$p(x) \sim \frac{c_1}{x^{1+\alpha_1}}, \quad x \rightarrow +\infty, \quad (6.12)$$

$$p(x) \sim \frac{c_2}{|x|^{1+\alpha_2}}, \quad x \rightarrow -\infty, \quad (6.13)$$

with $0 < \alpha_1, \alpha_2 \leq 2$. Then the parameter α is given by $\alpha = \min(\alpha_1, \alpha_2)$. The parameter β is related to the relative weight of the two tails of $p(x)$. If $\alpha_1 < \alpha_2$, the positive tail is dominant and $\beta = 1$. This is also what happens if one sums positive variables, that is, if $p(x) = 0$ for $x < 0$. More generally, this is the case when $p(-x)/p(x) \rightarrow 0$ when $x \rightarrow +\infty$. Conversely, if the negative tail is dominant ($\alpha_2 > \alpha_1$, and more generally $p(-x)/p(x) \rightarrow 0$ when $x \rightarrow -\infty$), one finds $\beta = -1$. In cases where both tails have comparable weights, that is, $\alpha_1 = \alpha_2$ so that $p(-x)/p(x)$ goes to a finite value when $x \rightarrow \infty$, the parameter β satisfies $-1 < \beta < 1$, and is given by

$$\beta = \frac{c_1 - c_2}{c_1 + c_2} \quad (6.14)$$

where c_1 and c_2 are the prefactors of the power-law decays of the tails, as given in Eqs. (6.12) and (6.13).

The choice of the parameters a_N and b_N also depends on α . When $\alpha > 1$, the mean value $\langle x \rangle$ is finite, so that a_N is simply equal to $N\langle x \rangle$ (with possibly an additive constant). When $\alpha \leq 1$, $\langle x \rangle$ is infinite, so that there is no point in characterizing the fluctuations around the mean. One then takes $a_N = 0$ (or again, possibly a constant non-zero value) for $\alpha < 1$ (the case $\alpha = 1$ involves logarithmic corrections). The Generalized Central Limit Theorem may then more naturally be interpreted, for $\alpha \leq 1$, as a generalization of the Law of Large Numbers. On the other side, the coefficient b_N formally takes a similar expression, namely $b_N \propto N^{1/\alpha}$, for all values of α ($0 < \alpha < 2$). The limit case $\alpha = 2$ corresponds to the standard rescaling used in the Central Limit Theorem. Note that for $\alpha = 2$, the Lévy distribution given in Eq. (6.8) identifies with the Gaussian distribution for all values of β .

6.1.3 Case of Non-identically Distributed Variables

In the previous subsection, we focused on the simplest case of independent and identically distributed random variables. As we already emphasized, this is a strong assumption whose validity can be questioned in many applications. Going beyond this assumption requires to consider either correlated variables (strictly speaking, dependent variables), or non-identically distributed variables. Of course, in a general situation, the random variables would be both correlated and non-identically distributed. We start by considering the case of independent, but non-identically distributed random variables. By definition, the probability distribution of such random variables factorizes into a product of non-identical functions of the individual variables (called the marginal distributions):

$$P_N(x_1, \dots, x_N) = \prod_{j=1}^N p_j(x_j). \quad (6.15)$$

Such a factorization property makes the analytical treatment easier, so that some of the results obtained for independent and identically distributed random variables can tentatively be generalized in the present framework. In particular, a generalized form of the Central Limit Theorem exists, if the following condition, called the Lindeberg condition [1], is satisfied. Consider a set (x_1, \dots, x_N) of N independent random variables with probability distribution $p_j(x)$, $j = 1, \dots, N$, with finite first and second moments. We denote as $m_j \equiv \langle x_j \rangle$ the first moment of x_j , and as $\sigma_j^2 \equiv \langle x_j^2 \rangle - m_j^2$ its variance. We further introduce the rescaling parameters

$$a_N = \sum_{j=1}^N m_j, \quad b_N = \left(\sum_{j=1}^N \sigma_j^2 \right)^{\frac{1}{2}}, \quad (6.16)$$

as well as the rescaled sum

$$z_N = \frac{1}{b_N} \left(\sum_{j=1}^N x_j - a_N \right), \quad (6.17)$$

which has a distribution $\Psi_N(z)$. Note that by definition of the rescaling parameters a_N and b_N , the variable z_N has zero mean and unit variance. The distribution $\Psi_N(z)$ converges to a Gaussian distribution if and only if the Lindeberg condition is satisfied, namely

$$\lim_{N \rightarrow \infty} \frac{1}{b_N^2} \sum_{j=1}^N \int_{|v| > \epsilon b_N} dv v^2 p_j(v + \langle x_j \rangle) = 0 \quad (6.18)$$

for all $\epsilon > 0$. Intuitively, the Lindeberg condition means that the individual terms in the sum become infinitesimal with respect to the sum in the limit of an infinite number of terms. It is clear that the Lindeberg condition holds for independent and identically distributed random variables, in which case the condition simply reads

$$\lim_{N \rightarrow \infty} \frac{1}{b_1^2} \int_{|v| > \epsilon b_N} dv v^2 p(v + \langle x \rangle) = 0, \quad (6.19)$$

using $b_N = b_1 \sqrt{N}$ (a condition valid only for independent and identically distributed variables). The fact that $b_N \rightarrow \infty$ ensures that condition (6.19) is satisfied. Coming back to the general case of non-identically distributed variables, it is worth noticing that the Lindeberg condition implies that the variance b_N of the sum diverges when the number of terms goes to infinity. This can be checked by showing that if b_N is bounded, the Lindeberg condition cannot be satisfied. As b_N is by definition an increasing function of N , assuming that it is bounded implies that b_N goes to a finite limit when $N \rightarrow \infty$. As a result, all the integrals that are summed in Eq. (6.18) become independent of N (although they generically depend on j) in the large N limit. All these integrals are non-negative by definition. For small enough ϵ , the integral corresponding to $j = 1$ is strictly positive, so that the sum in Eq. (6.18) has to be larger than a strictly positive bound. Since the prefactor $1/b_N^2$ converges, by assumption, to a finite limit, the limit in Eq. (6.18) cannot be equal to zero.

In cases where the Lindeberg condition does not hold, $\Psi_N(z)$ converges to a non-Gaussian limit distribution, which depends on the specific problem at hand. As an explicit example of non-Gaussian distribution appearing in this context, one can mention the following simple $1/f$ -noise model [3]. In this model, one considers a random time signal $h(t)$ that is discretized into a sequence of values h_k , $k = 0, \dots, N - 1$. Note that we consider t as a time for the sake of simplicity, but t could alternatively be interpreted as a space coordinate in a one-dimensional system. The discretized signal h_k can be analyzed through a discrete Fourier transform, defining the complex Fourier amplitude

$$c_n = \frac{1}{\sqrt{N}} \sum_{\ell=0}^{N-1} h_\ell e^{-2i\pi f_n \ell} \quad (6.20)$$

associated with the frequency $f_n = n/N$, $n = 0, \dots, N-1$. As a simplification, the present $1/f$ -noise model assumes that the Fourier coefficients c_n are statistically independent random variables, with real and imaginary parts distributed according to Gaussian distributions of variance $\sigma_n^2 = \kappa/n$ for $n > 0$ (hence the name $1/f$ -noise, since the frequency is proportional to n). Introducing the average value $\bar{h} = N^{-1} \sum_{\ell=0}^{N-1} h_\ell$, the fluctuations of the signal are characterized by its empirical variance (or “roughness”)

$$E_N = \sum_{\ell=0}^{N-1} (h_\ell - \bar{h})^2, \quad (6.21)$$

which is also a random variable. Using Parseval’s theorem, the empirical variance may be rewritten as

$$E_N = \sum_{n=1}^{N-1} |c_n|^2. \quad (6.22)$$

In this way, E_N also appears as the total energy (or integrated spectrum) of the signal. Note that the Fourier coefficient c_0 , which is proportional to the average value of the signal, plays no role here and thus does not appear in Eq. (6.22), since we focus on fluctuations around the mean value. At this stage, we see that E_N turns out to be a sum of independent, but non-identically distributed variables $u_n \equiv |c_n|^2$. One can show, from the Gaussian distributions of the real and imaginary parts of c_n , that the distribution of u_n is exponential,

$$\tilde{p}_n(u_n) = n\kappa e^{-n\kappa u_n}. \quad (6.23)$$

It follows that the variance of u_n is given by κ^2/n^2 , according to which the variance $\text{Var}(E_N) = \sum_{n=1}^{N-1} \text{Var}(u_n)$ goes to a finite limit when $N \rightarrow \infty$. As a result, the Lindeberg condition does not hold, showing that the limit distribution is non-Gaussian.

In order to determine the limit distribution in the large N limit, we rescale the energy into $\varepsilon = (E_N - \langle E_N \rangle)/\sigma$, where σ^2 is the infinite N limit of the variance of E_N , that is $\sigma^2 = \sum_{n=1}^{\infty} \kappa^2/n^2$. The distribution of ε_N is denoted as $\Psi_N(\varepsilon)$. In order to determine $\Psi_N(\varepsilon)$, it is convenient to define the characteristic function

$$\chi_N(\omega) = \int_{-\infty}^{\infty} d\varepsilon \Psi_N(\varepsilon) e^{-i\omega\varepsilon}. \quad (6.24)$$

Since the variables u_n are independent, the characteristic function of the sum is simply the product of the characteristic functions of the variables u_n , $n = 1, \dots, N-1$. Taking the limit $N \rightarrow \infty$, the characteristic function $\chi_N(\omega)$ converges to a limit function $\chi_\infty(\omega)$, which reads [3, 4]

$$\chi_\infty(\omega) = \prod_{n=1}^{\infty} \left(1 + \frac{i\omega}{n\sigma\kappa}\right)^{-1} \exp\left(\frac{i\omega}{n\sigma\kappa}\right). \quad (6.25)$$

The expression can be put into a more manageable form using the relation

$$\Gamma(1+z) = e^{-\gamma_E z} \prod_{n=1}^{\infty} \frac{e^{z/n}}{1 + \frac{z}{n}}, \quad (6.26)$$

with $\gamma_E = 0.577\dots$ the Euler constant, and where z is an arbitrary complex number satisfying $z \neq -1, -2, \dots$ [5]. The Euler Gamma function Γ appearing in Eq. (6.26) is defined as $\Gamma(x) = \int_0^\infty dt t^{x-1} e^{-t}$. The inverse Fourier transform of $\chi_\infty(\omega)$ can be computed explicitly, yielding a Gumbel distribution

$$\Psi_\infty(\varepsilon) = \exp[-(b\varepsilon + \gamma_E) - e^{-(b\varepsilon + \gamma_E)}], \quad b = \frac{\pi}{\sqrt{6}}. \quad (6.27)$$

This is a rather surprising result as the Gumbel distribution is known to appear usually in the context of extreme value statistics, as described below. The fact that it also appears in the present model which bears no obvious relation to extreme value statistics can actually be understood in a simple way [4, 6], which we however do not detail here.

One of the main interests of the present $1/f$ -noise model is to illustrate the fact that correlated random variables (the original signal h_k) can in some cases be converted, through a Fourier transform, into a sequence of independent, but non-identically distributed variables (the amplitudes c_n), which allows for a simpler analytical treatment. Of course, the assumption that the Fourier amplitudes are statistically independent random variables is an approximation with respect to realistic systems, but it already captures the onset of a non-Gaussian limit distribution for the total energy, which is a result of interest. In more general cases however, a problem of sum of correlated random variables cannot be so easily converted into a problem of independent random variables, and one has to deal directly with the correlated case.

6.1.4 Case of Correlated Variables

Determining the limit distribution of the sum of a sequence of correlated variables may be a difficult task. There exist, however, results for specific classes of correlated random variables, like martingale differences [1], or functionals of stationary Gaussian sequences [7, 8]. On a less rigorous basis, arguments have also been proposed in the physics literature to generalize the Central Limit Theorem to some classes of correlated random variables, for instance by considering “deformed products” [9] or related notions based on the non-extensive entropy formalism [10]. Results based on a class of random variables with a joint probability expressed as

a product of matrix functions (instead of real functions in the case of independent random variables) have also been proposed recently [11].

From a heuristic point of view, the emergence of non-Gaussian distributions in strongly correlated variables can be understood as follows. Let us consider a large system in which microscopic degrees of freedom are correlated over a typical length ξ , smaller than the system size L . One can then virtually decompose the system into boxes of linear size of the order of ξ , so that correlations between different boxes are weak. In many cases, the global variable of interest can be decomposed as a sum of contributions from each box. Then, the global observable can be approximated as a sum of $N = (L/\xi)^D$ independent random variables, where N is the number of boxes (and D is the space dimension). Assuming that the local observable in each box has a finite variance, the distribution of the sum tends to a Gaussian distribution when the number of boxes N goes to infinity. This is the case for instance if the correlation length ξ is fixed and the system size L goes to infinity. Yet, it happens in some physical systems that the correlation length is proportional to the system size L , so that ξ/L takes a finite value when $L \rightarrow \infty$ (this is the case for instance in generalizations of the $1/f$ -noise problem [4]). In this situation, the number of independent boxes remains finite, and the Central Limit Theorem does not hold, so that the limit distribution of the sum is not Gaussian.

Beyond this type of heuristic arguments, some rigorous results exist in particular for Gaussian sequences of random variables. Such sequences are defined by the following joint probability distribution

$$P_N(x_1, \dots, x_N) = \frac{\sqrt{\det R}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2} \sum_{i,j=1}^N x_i R_{ij}^{-1} x_j\right) \quad (6.28)$$

where R is a positive-definite matrix of determinant $\det R$, and R_{ij}^{-1} are the elements of the inverse matrix R^{-1} . Note that for simplicity, we focus here on the case of centered variables, for which $\langle x_i \rangle = 0$. Generalization to non-centered variables is however straightforward. If there exists a function $r(m)$ such that the matrix elements of R satisfy $R_{ij} = r(|i - j|)$, the Gaussian sequence is said to be stationary. In this case, the marginal distribution of x_i is a centered Gaussian distribution of variance $r(0)$, while the two-point correlation $\langle x_i x_{i+m} \rangle$ is equal to $r(m)$.

If a large Gaussian stationary sequences (x_1, \dots, x_N) is characterized by a correlation function $r(m)$ with a power-law decay at large distance, $r(m) \sim m^{-\alpha}$ ($\alpha > 0$), then the distribution of the sum $S_N = \sum_{i=1}^N x_i$ converges (as usual, up to a rescaling $z_N = (S_N - a_N)/b_N$) to a Gaussian distribution for all values of $\alpha > 0$ [7, 8]. Hence for Gaussian sequences, even very strong correlations do not prevent the distribution of the sum from converging to a Gaussian limit. For stationary sequences of non-Gaussian correlated variables, a Gaussian distribution of the sum is obtained at least when the sum $\sum_{m=1}^n r(m)$ converges for $n \rightarrow \infty$ [12].

Deviations from the Gaussian limit distribution may be obtained by considering non-Gaussian sequences of random variables, with strong enough correlations. Generic results exist for variables obtained as nonlinear transforms $y_i = \psi(x_i)$ of the Gaussian variables x_i [7, 8]. These results rely on the expansion of the function $\psi(x)$ onto the basis of Hermite polynomials. The detailed presentation of this result goes beyond the scope of the present book. However, a simple application of the theorem can be provided. In the case where $y_i = x_i^2 - 1$ [13], the theorem states that the limit distribution of the rescaled sum z_N is Gaussian as long as $\alpha > 1/2$, while it is non-Gaussian when $\alpha < 1/2$. The corresponding non-Gaussian limit distribution is known through its cumulant expansion.

6.1.5 Coarse-Graining Procedures and Law of Large Numbers

As mentioned above, the Law of Large Numbers and the Central Limit Theorem have many applications in statistical physics. We have seen for instance in Sect. 2.1.3 direct applications of the Central Limit Theorem in the context of random walks. Standard random walks are described by the standard Central Limit Theorem, while random walks with broad distributions of jump size (i.e., with infinite variance) are described by the Generalized Central Limit Theorem. Accordingly, the distribution of the position of the random walk is Gaussian in the first case, and is a Lévy distribution in the second case.

Here, we would like to briefly address the role of the Law of Large Numbers in the derivation of large scale equations describing continuous fields like the density field. Let us consider a model similar to the Zero Range Process, though we do not explicitly specify the dynamics. The model is defined on a one-dimensional lattice with L sites, and describes identical particles hopping between different sites. The number of particles on site $i = 1, \dots, L$ is denoted as n_i . In order to coarse-grain the model, we further split the lattice into boxes containing a number ℓ of sites, such that $1 \ll \ell \ll L$. A given box is labelled by an almost continuous variable $x = i_0/L$, where i_0 is the site at the center of the box. The width of the boxes is equal to $\Delta x = 1/L$. We then define a coarse-grained density field as

$$\rho(x) = \frac{1}{\Delta x} \sum_{i \in B(x)} n_i. \quad (6.29)$$

Turning to dynamics, the model is defined in such a way that particles can be exchanged from any site within a given box to any site within neighboring boxes. Given a time interval $[t, t + \Delta t]$, we introduce the number $\phi_{i,j}(t, \Delta t)$ of particles transferred between i and j (with i and j belonging to different boxes) in this interval,

counted positively from i to j and negatively from j to i ($i < j$). Coarse-graining this particle transfer at the level of boxes leads one to introduce the quantity

$$Q(x, t, \Delta t) = \sum_{i \in B(x), j \in B(x+\Delta x)} \phi_{i,j}(t, \Delta t). \quad (6.30)$$

The balance of particle transfers then reads

$$\rho(x, t + \Delta t) - \rho(x, t) = \frac{1}{\Delta x} [Q(x - \Delta x, t, \Delta t) - Q(x, t, \Delta t)]. \quad (6.31)$$

We now assume that the random variables n_i corresponding to different sites (whether in the same box or not) are statistically independent, with the same distribution—this property is true for instance in the Zero Range Process, assuming it is in contact with a reservoir of particles. We further assume that the statistical independence property is still valid when the system is close to a stationary state. As a result, the Law of Large Numbers can be applied, in the large box size limit, to the density of particles in each box since it is defined as a sum of a large number of independent and identically distributed variables. Hence it follows that the random variable $\rho(x)$ can be identified with its ensemble average value $\bar{\rho}(x)$.

In the same way, and under a similar assumption of statistical independence, the coarse-grained transfer of particles $Q(x, t, \Delta t)$ can be identified with its ensemble average value $\bar{Q}(x, t, \Delta t)$. This allows for further simplifications, since $\bar{Q}(x, t, \Delta t)$ is, for sufficiently small Δt , proportional to Δt :

$$\bar{Q}(x, t, \Delta t) = J(x, t) \Delta t. \quad (6.32)$$

Expanding Eq. (6.31) to first order in Δt and Δx , we eventually get

$$\frac{\partial \bar{\rho}}{\partial t}(x, t) = -\frac{\partial J}{\partial x}(x, t) \quad (6.33)$$

which is the general form of the continuity equation for the density field, in one dimension—see also Eq. (349) for the two-dimensional version in the case of self-propelled particles. The above reasoning illustrates how the Law of Large Numbers can be of key importance in order to justify coarse-graining procedures into continuous fields obeying deterministic equations like Eq. (6.33). To conclude this discussion, two remarks are in order. First, the above model was build in an ad hoc way, to fulfill all requirements in order to safely apply the Law of Large Numbers. One generally faces several difficulties when trying to coarse-grain more generic models: the random variables may not be independent and identically distributed, and the flux of particles between boxes may not be a sum of many contributions, but rather a small number of local terms through the box boundaries. In this case, more sophisticated methods need to be used to perform the coarse-graining.

Second, it is worth noticing that one can go beyond the Law of Large Numbers and try to describe the tiny fluctuations of the coarse-grained fields using the Central Limit Theorem. This is done in practice by adding a Gaussian noise term to the flux J [14]. The density field then remains a stochastic variable, but it is possible to decompose it into an average value plus some fluctuations whose evolution can be characterized through the continuity equation.

6.2 Rare and Extreme Events

6.2.1 *Different Types of Rare Events*

In full generality, rare events are simply events with a low probability (one sometimes says “in the tail of the probability distribution”). Yet, to deserve interest, what is often called a rare event is implicitly an event that, on top of occurring unfrequently, may also have a significant impact on the evolution of the system, if the event is relative to some observable defined on a given system.

Beyond these very generic characterizations, the notion of rare event may actually refer to several types of events. A first type of rare event is that associated to the crossing of a threshold or an energy barrier for instance, like in chemical reaction pathways or during a local rearrangement of a dense amorphous material. If the threshold or barrier is high, the typical time to overcome it is very large, and the crossing event thus has a very low probability. Yet, this crossing is key to the dynamics, and allows the systems to explore different regions of phase space, thereby allowing the chemical reaction between molecules to occur, or the material to deform and to relax its stress. Hence this is typically a situation in which a rare event has a noticeable impact on the dynamics of the system.

A second type of rare event corresponds to extreme values, that is to maximum or minimum values in a set of random variables or in practice empirical data, and the related notion of records. These important notions are respectively the subjects of the next two Sects. 6.2.2 and 6.2.3. Moreover, a third type of rare event may be identified as the effect of extreme events on a sum of broadly distributed random variables, as we have seen in Chap. 2 in the case of anomalous diffusion, and in the present chapter when discussing the Generalized Central Limit Theorem in Sect. 6.1.2.

Finally, a last type of rare event corresponds to extremely rare events, that cannot be observed in practice unless a control parameter of the system is tuned to make them become typical. This situation is one of the basic motivations to introduce the notion of large deviation function, that characterizes events that have a probability which decreases exponentially with system size (typically its volume). This is the topic of Sect. 6.3.

6.2.2 Extreme Value Statistics

In this section, we provide some standard results on the statistics of extreme values, namely the maximum or minimum value in a set of random variables. For simplicity, it is convenient to restrict the study to the case of maximum values, as the case of minimum values can be mapped onto the one of maximum values through a change of sign. Let us consider a sequence (x_1, \dots, x_N) of independent and identically distributed random variables, with probability distribution $p(x)$, called the parent distribution. It is useful to also introduce the cumulative distribution, defined as

$$F(x) = \int_{-\infty}^x p(x') dx' \quad (6.34)$$

which is the probability that the random variable is smaller than a value x . For a given sequence of variables (x_1, \dots, x_N) , one can define the maximum value in the set,

$$y_N = \max(x_1, \dots, x_N). \quad (6.35)$$

We denote as $F_N^{\max}(y)$ the probability that the maximum y_N is smaller than a given value y . By definition of the maximum, $x_i \leq y_N$ for all i . Hence $F_N^{\max}(y)$ is equal to the probability that all the variables x_i are smaller than y . Since we are considering independent and identically distributed random variables, $F_N^{\max}(y)$ can simply be written as

$$F_N^{\max}(y) = F(y)^N. \quad (6.36)$$

In the same spirit as one may look for the limit distribution of the rescaled sum in the context of the Central Limit Theorem, one is interested here in the limit distribution of the rescaled maximum of the set (x_1, \dots, x_N) . Introducing a sequence of rescaling parameters a_N and b_N , we first define the rescaled maximum as

$$z_N = \frac{y_N - a_N}{b_N}. \quad (6.37)$$

The question is then to know whether, for a suitable choice of a_N and b_N , the cumulative distribution $H_N(z)$ of the rescaled maximum z_N converges to a limit distribution.

Standard results of extreme value statistics [15, 16] lead to the following statement. Depending on the distribution $p(x)$ of the variables x_i , the limit cumulative distribution, $\lim_{N \rightarrow \infty} H_N(z)$, can take three different forms:

- $H_f(z) = \exp(-z^{-\mu})$ for $z > 0$ and 0 for $z \leq 0$ (Fréchet distribution);
- $H_w(z) = \exp(-(-z)^\mu)$ for $z < 0$ and 1 for $z \geq 0$ (Weibull distribution);
- $H_g(z) = \exp(-e^{-z})$ (Fisher-Tippett-Gumbel or Gumbel distribution).

In these distributions, μ is a positive parameter, related to the parent distribution $p(x)$. The probability densities $p_f(z)$, $p_w(z)$ and $p_g(z)$ are obtained from the cumulative

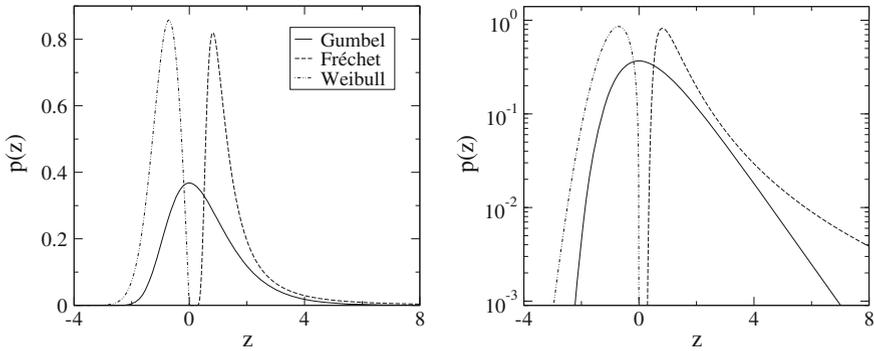


Fig. 6.2 *Left* Illustration of the Gumbel (*full line*), Fréchet (*dashed line*) and Weibull (*dot-dashed*) probability densities. Both the Fréchet and Weibull distributions have a parameter $\mu = 2$. *Right* Same data on semi-logarithmic scale

distributions $H_f(z)$, $H_w(z)$ and $H_g(z)$ by taking the derivative. As an example, the Gumbel probability density is given by

$$p_g(z) = \exp(-z - e^{-z}). \tag{6.38}$$

An illustration of the Fréchet, Weibull and Gumbel distributions is provided in Fig. 6.2. Note that the cumulative distributions $H_f(z)$, $H_w(z)$ and $H_g(z)$ can also be formulated in a compact way using a single expression

$$H_\gamma(z') = \exp(-(1 + \gamma z')^{-1/\gamma}), \quad 1 + \gamma z' > 0 \tag{6.39}$$

where $z' = az + b$, a and b being some constant parameters which depend on γ . The case $\gamma > 0$ corresponds to the Fréchet distribution, with the relation $\mu = 1/\gamma$. The case $\gamma < 0$ instead corresponds to the Weibull distribution, for which $\mu = -1/\gamma$. Finally, the case $\gamma = 0$, to be interpreted as the limit $\gamma \rightarrow 0$ in Eq. (6.39), corresponds to the Gumbel distribution.

The different distributions $H_f(z)$, $H_w(z)$ and $H_g(z)$ are selected according to some asymptotic large x properties of the parent distribution $p(x)$. If $p(x)$ decays as a power law $p(x) \sim 1/x^{1+\mu}$ when $x \rightarrow \infty$ ($\mu > 0$), the limit distribution is the Fréchet one with the same exponent μ as the one characterizing the parent distribution $p(x)$. If rather the variable x is bounded by a constant A in the sense that $p(x) = 0$ for $x > A$, and $p(x)$ behaves as a power law close to $x = A$, namely $p(x) \sim (A - x)^{\mu-1}$ with $\mu > 0$, the limit distribution is the Weibull one. Finally, in the case where the distribution $p(x)$ decays faster than any power law, either for x going to infinity or for x going to a finite bound A , the limit distribution is the Gumbel one. It is customary to say that the distribution $p(x)$ belongs either to the Fréchet, Weibull or Gumbel class according to the limit distribution of the maximum. A typical example of a distribution $p(x)$ belonging to the Gumbel class is the exponential distribution $p(x) = \lambda \exp(-\lambda x)$.

Let us illustrate in this simple case how the limit distribution can be derived. The corresponding cumulative probability distribution is $F(x) = 1 - \exp(-\lambda x)$ for $x > 0$ and $F(x) = 0$ for $x \leq 0$. Then the cumulative probability distribution $F_N^{\max}(y)$ of the maximum is given (for $y > 0$) by

$$F_N^{\max}(y) = F(y)^N = (1 - e^{-\lambda y})^N. \quad (6.40)$$

The goal is to find rescaling parameters a_N and b_N such that the distribution of the rescaled maximum $z_N = (y_N - a_N)/b_N$ converges to a limit distribution, in this case the Gumbel distribution. Considering Eq. (6.40), one finds by inspection that a correct rescaling is obtained by choosing $a_N = (\ln N)/\lambda$ and $b_N = 1/\lambda$. Inverting the definition of z_N to get $y_N = a_N + b_N z$, we have the following convergence property:

$$F_N^{\max}(a_N + b_N z) = \left(1 - \frac{e^{-z}}{N}\right)^N \rightarrow \exp(-e^{-z}) \quad (6.41)$$

when $N \rightarrow \infty$. The derivation of the Gumbel limit distribution starting from an arbitrary cumulative distribution $F(x)$ is more complicated, but the spirit of the derivation remains the same.

6.2.3 Statistics of Records

A notion closely related to that of extreme value is that of record. A record in a sequence of random variables $(x_1, x_2, \dots, x_i, \dots)$ occurs at the n th step when the value x_n is larger than all previous values x_i , $i = 1, \dots, n - 1$ (for simplicity, we focus here on upper records; lower records are obtained symmetrically as minimal values). Of course, the record value x_n is also the maximum value of the set of variables (x_1, \dots, x_n) , but the questions asked in extreme value statistics and in record statistics are slightly different. In extreme value statistics, one considers a fixed number N of variables, (x_1, \dots, x_N) , and asks about the statistics of the maximum value $y_N = \max(x_1, \dots, x_N)$ in the set; the limit $N \rightarrow \infty$ is eventually taken. Note that the order of the variables in the set (x_1, \dots, x_N) plays no role. In record statistics, one rather looks at the occurrence of successive records, so that the order of variables in a given sample matters, and the sequence does not have a fixed length, but is rather considered to be infinite from the outset. One thus defines the k th record in a recursive way. The first variable x_1 defines the first record r_1 . Then one looks at the next variables (x_2, x_3, \dots) in the sequence for the occurrence of the second record r_2 , that is, the first variable x_j ($j > 1$) such that $x_j > r_1$. This occurs for a value $j = n_2$, and we have $r_2 = x_{n_2}$. In the same way, one defines recursively the record $r_k = x_{n_k}$, which is the first variable in the sequence exceeding the previous record r_{k-1} .

There are typically two types of quantities that can be investigated in the framework of record statistics: first, the statistics of the “time” n_k at which the k th record

occurs; second, the statistics of the records r_k themselves. In the latter case, one may be interested in looking for the limit distribution of the variable r_k in the limit $k \rightarrow \infty$, up to a suitable rescaling as done in the case of extreme value statistics.

We will consider here only the most basic statistical properties of n_k . Interestingly, for independent and identically distributed random variables x_i , these properties do not depend on the probability distribution $p(x)$, but are universal [15, 17]. Let us start by considering the probability P_n that the n th variable in the sequence, x_n , is a record. This probability reads

$$P_n = \int_{-\infty}^{\infty} p(x_n) F(x_n)^{n-1} dx_n \quad (6.42)$$

where $F(x) = \int_{-\infty}^x p(x) dx$ is the cumulative distribution of x . Eq. (6.42) is simply obtained by averaging over x_n the probability $F(x_n)^{n-1}$ that the $n-1$ other variables x_1, \dots, x_{n-1} are smaller than x_n . Noting that

$$\frac{d}{dx_n} F(x_n)^n = np(x_n)F(x_n)^{n-1} \quad (6.43)$$

we easily obtain that $P_n = \frac{1}{n}$, independently of the distribution $p(x)$. An immediate consequence is that the average number N_n of records occurring up to “time” n is given by

$$N_n = \sum_{k=1}^n P_k = \sum_{k=1}^n \frac{1}{k} \quad (6.44)$$

which in the large n limit behaves logarithmically to leading order,

$$N_n \approx \ln n + \gamma_E \quad (6.45)$$

where $\gamma_E \approx 0.577$ is the Euler constant.

In contrast, the asymptotic limit distribution of the records r_k depends on the distribution $p(x)$ of the variables in the sequence, but only through classes of limit distributions, similarly to the case of extreme value statistics. One can here again split the distributions $p(x)$ into three different classes, namely Gumbel, Fréchet and Weibull, depending on their asymptotic behavior, which allows one to define the parameter γ in a similar way as in extreme value statistics, see Eq. (6.39). As above, the Gumbel class corresponds to distributions $p(x)$ decaying faster than any power law (typically exponentially), the Fréchet class to distributions decaying as a power law at infinity, while the Weibull class describes distributions behaving as a power law close to an upper bound. The limit distributions are however different from that obtained in extreme value statistics. Let us introduce the rescaled k th record

$$z_k = \frac{r_k - a_k}{b_k} \quad (6.46)$$

where a_k and b_k are suitably chosen rescaling parameters. We denote as $R_k(z)$ the cumulative distribution of z_k . If the distribution $p(x)$ is in the Gumbel class ($\gamma = 0$), the limit distribution $\lim_{k \rightarrow \infty} R_k(z)$ is given by [15]

$$R_g(z) = \Phi(z) \quad (6.47)$$

where $\Phi(z)$ is the integrated normal distribution,

$$\Phi(z) = \int_{-\infty}^z e^{-y^2/2} dy. \quad (6.48)$$

In other words, the limit distribution is simply a Gaussian (or normal) distribution. For the Fréchet class ($\gamma > 0$), the limit distribution reads

$$R_f(z) = \Phi(\gamma \ln x), \quad x > 0 \quad (6.49)$$

thus corresponding to a (positive) lognormal distribution. Conversely, for the Weibull class ($\gamma < 0$), the limit distribution is given by

$$R_w(z) = \Phi(\gamma \ln(-x)), \quad x < 0 \quad (6.50)$$

which corresponds to a (negative) lognormal distribution.

6.3 Large Deviation Functions

We have already encountered the notion of large deviation form of a probability distribution, for instance in the case of phase transitions (Sect. 1.4), reaction-diffusion processes (Sect. 4.1), or random networks (Sect. 4.3). However, this form only appeared as a formal property in these previous examples, and we wish to discuss here the interest and interpretation of such a form.

6.3.1 A Simple Example: The Ising Model in a Magnetic Field

To illustrate the notion of large deviation function and the relevance to describe extremely rare events, let us consider a simple example, the effect of a magnetic field h on an Ising model at high temperature, well above the ferromagnetic transition

temperature. In this case, the coupling energy between spins can be safely neglected with respect to the thermal energy (i.e., $J/T \ll 1$). Let us first consider the case of a zero magnetic field ($h = 0$). By simply counting the configurations having a macroscopic magnetization

$$m = \frac{1}{N} \sum_{i=1}^N s_i \quad (6.51)$$

where N is the number of spins, one obtains for the probability distribution $P(m)$

$$P(m, h = 0) \propto e^{-Ncm^2} \quad (6.52)$$

for not too large values of m , with some constant $c > 0$ [see Eqs. (1.92) and (1.93)].

Hence any value $m \neq 0$ has a vanishingly small probability to be observed in the thermodynamic limit $N \rightarrow \infty$. Equation (6.52) is a simple example of large deviation form of a distribution. More generally, a distribution function $P(x)$ has a large deviation form if it takes the asymptotic form, for large N ,

$$P(x) \propto e^{-N\phi(x)} \quad (6.53)$$

where $\phi(x)$ is called the large deviation function, or rate function. A more rigorous definition can be written as

$$\phi(x) = - \lim_{N \rightarrow \infty} \frac{1}{N} \ln P(x). \quad (6.54)$$

In this general setting, N may be the number of particles, of spins, or the volume of the system. In the case of the paramagnetic model, Eq. (6.52) yields for the large deviation function $\phi(m, h = 0) = cm^2$.

In the presence of a magnetic field $h \neq 0$, one then finds

$$P(m) \propto e^{-Ncm^2 + Nhm/kT} \propto e^{-Nc(m-m_0)^2}, \quad m_0 = \frac{h}{2ckT} \quad (6.55)$$

or in other words $\phi(m, h) = c(m-m_0)^2$. Hence in the presence of a magnetic field, the magnetization m_0 , which was extremely rare and in practice unobserved for $h = 0$, becomes the typical value. Varying an external control parameter thus makes typical a value of the observable that was extremely rare otherwise. The interest of the notion of large deviation function therefore partly resides in this property. Characterizing the extremely low probability of a random variable is not so interesting in itself: whether the probability of a given event is 10^{-40} or 10^{-100} does not make much difference, as the event will never be observed in practice. However, knowing this very low probability enables one to predict the effect of an external control parameter like a magnetic field, which acts as a simple exponential reweighting of the zero field probability:

$$P(m, h) \propto P(m, 0) e^{Nhm} \quad (6.56)$$

A review of the use of large deviation functions in a statistical physics context can be found in Ref. [18].

6.3.2 *Explicit Computations of Large Deviation Functions*

Large deviations functions can be computed thanks to the Gärtner-Ellis theorem, which can be (loosely) stated as follows [18]. Given a set of random variables x_N indexed by an integer N and defined over an interval (a, b) , the distribution $p_N(x)$ takes a large deviation form

$$p_N(x) \propto e^{-N\phi(x)} \quad (6.57)$$

if the following scaled-cumulant generating function

$$\lambda(k) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \langle e^{Nkx_N} \rangle, \quad (6.58)$$

with k real, exists (i.e., takes finite values) over some interval of k , possibly the whole real axis. Then the large deviation function exists and is given by the Legendre-Fenchel transform of $\lambda(k)$,

$$\phi(x) = \sup_k [kx - \lambda(k)]. \quad (6.59)$$

At a heuristic level, this relation can be understood as follows, assuming the validity of the large deviation form Eq. (6.57). To compute $\lambda(k)$, one first needs to evaluate

$$\langle e^{Nkx_N} \rangle = \int_a^b dx e^{N[kx - \phi(x)]}. \quad (6.60)$$

When k is such that the maximum x_k^* of $kx - \phi(x)$ falls within the interval (a, b) , the integral can be evaluated in the large N limit through a saddle-point approximation,

$$\int_a^b dx e^{N[kx - \phi(x)]} \sim e^{N[kx_k^* - \phi(x_k^*)]} \quad (6.61)$$

leading to

$$\lambda(k) = kx_k^* - \phi(x_k^*) = \sup_{x \in (a,b)} [kx - \phi(x)]. \quad (6.62)$$

Hence $\lambda(k)$ is the Legendre-Fenchel transform of $\phi(x)$. Inverting this transform precisely yields Eq. (6.59).

A simple application of this theorem is provided by the case of the sum of independent and identically distributed random variables (u_1, \dots, u_N) of distribution $P(u)$. Defining x_N as the empirical mean of the variables u_i ,

$$x_N = \frac{1}{N} \sum_{i=1}^N u_i, \quad (6.63)$$

we can test whether the distribution $p_N(x)$ of x_N takes a large deviation form. Following the Gärtner-Ellis theorem, we compute $\lambda(k)$, yielding

$$\lambda(k) = \ln \langle e^{ku} \rangle \quad (6.64)$$

where the brackets here mean an average over the distribution $P(u)$. The large deviation function is then obtained by Eq. (6.59). For example, for an exponential distribution $P(u) = e^{-u}$, we have $\lambda(k) = -\ln(1 - k)$ and thus $\phi(x) = x - 1 - \ln x$ ($x > 0$).

6.3.3 A Natural Framework to Formulate Statistical Physics

Large deviation functions turn out to be a natural language for statistical physics, as can be already seen at equilibrium. We have seen in particular when studying equilibrium phase transitions that the distribution of magnetization in the mean-field Ising model takes a large deviation form

$$P(m) \propto e^{-Nf(m)} \quad (6.65)$$

where $f(m)$ is given in Eq. (1.93). This function has been seen to provide useful information on the phase transition. This is actually another example of the usefulness of large deviation functions. In this mean-field case, the computation of the large deviation function is easy (which is not the case in general as soon as there are correlations—or interactions—in the system), thus providing a direct characterization of the phase transition. Hence determining the whole probability distribution of events that are for most of them unobservable is actually one of the easiest ways to compute the physically observed values. This also has the further advantage to predict the two symmetric most probable values of the magnetization, while a direct computation of the mean magnetization would result in an average over the two symmetric values, hence to $m = 0$.

The importance of large deviation functions in equilibrium statistical physics also comes from the fact that basic quantities like the phase space volume $\Omega_N(E)$ or partition functions $Z_N(T)$ take large deviation forms

$$\Omega_N(E) \propto e^{Ns(\varepsilon)}, \quad Z_N(T) \propto e^{-Nf(T)/kT} \quad (6.66)$$

showing that the entropy per degree of freedom $s(\varepsilon)$ (with $\varepsilon = E/N$) and the (rescaled) free energy $f(T)/kT$ play the role of large deviation functions (although in a less restricted sense than that previously introduced, since $\Omega_N(E)$ and $Z_N(T)$ are not probability distributions).

Turning to out-of-equilibrium situations, we have seen an example of the use of a large deviation function in a nonequilibrium context when discussing absorbing phase transitions as well as networks—see Chap. 4. More generally, there has been several attempts to use large deviation functions in nonequilibrium models in order to generalize the equilibrium notion of free energy [19, 20]. Such attempts however go much beyond the scope of the present book, and will not be discussed here.

References

1. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd edn. (Wiley, New York, 1971)
2. B.V. Gnedenko and A.N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley, 1954)
3. T. Antal, M. Droz, G. Györgyi, Z. Rácz, $1/f$ noise and extreme value statistics. *Phys. Rev. Lett.* **87**, 240601 (2001)
4. E. Bertin, M. Clusel, Generalised extreme value statistics and sum of correlated variables. *J. Phys. A* **39**, 7607 (2006)
5. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, 5th edn. (Academic Press, London, 1994)
6. E. Bertin, Global fluctuations and Gumbel statistics. *Phys. Rev. Lett.* **95**, 170601 (2005)
7. M.S. Taqqu, Convergence of iterated process of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 53 (1979)
8. P. Breuer, P. Major, Central limit theorems for non-linear functionals of Gaussian fields. *J. Multivar. Anal.* **13**, 425 (1983)
9. F. Baldovin, A.L. Stella, Central limit theorem for anomalous scaling due to correlations. *Phys. Rev. E* **75**, 020101(R) (2007)
10. S. Umarov, C. Tsallis, S. Steinberg, On a q-Central limit theorem consistent with nonextensive statistical mechanics. *Milan J. Math.* **76**, 307 (2008)
11. F. Angeletti, E. Bertin, P. Abry, General limit distributions for sums of random variables with a matrix product representation. *J. Stat. Phys.* **157**, 1255 (2014)
12. J.-P. Bouchaud, A. Georges, Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Phys. Rep.* **195**, 127 (1990)
13. M. Rosenblatt, Independence and dependence, in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2 (University of California, 1961), p. 431
14. J.M. Ortiz de Zarate, J.V. Sengers, *Hydrodynamic Fluctuations in Fluids and Fluid Mixtures* (Elsevier, 2006)
15. J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (Wiley, New York, 1987)

16. E.J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958; Dover Publication, 2004)
17. J. Krug, Records in a changing world. *J. Stat. Mech.*, P07001 (2007)
18. H. Touchette, The large deviation approach to statistical mechanics. *Phys. Rep.* **478**, 1 (2009)
19. B. Derrida, J.L. Lebowitz, E.R. Speer, Free energy functional for nonequilibrium systems: an exactly solvable case. *Phys. Rev. Lett.* **87**, 150601 (2001)
20. B. Derrida, J.L. Lebowitz, E.R. Speer, Exact free energy functional for a driven diffusive open stationary nonequilibrium system. *Phys. Rev. Lett.* **89**, 030601 (2002)

Appendix A

Dirac Distribution

The Dirac distribution $\delta(x)$ can be thought of as a function being equal to zero for all $x \neq 0$, and being infinite for $x = 0$, in such a way that $\int_{-\infty}^{\infty} \delta(x) dx = 1$. The main interest of the Dirac distribution is that for an arbitrary function f ,

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_0) dx = f(x_0) \tag{A.1}$$

where x_0 is an arbitrary constant. In other words, once inserted in an integral, the Dirac distribution precisely picks up the value of the integrand associated to the value of the variable around which it is peaked.

The following property, related to changes of variables in the calculation of integrals, also proves useful. Suppose one needs to compute the integral

$$I(a) = \int_{x_{\min}}^{x_{\max}} dx g(x) \delta(f(x) - a) \tag{A.2}$$

where $g(x)$ is an arbitrary function. Such integrals appear for instance in the computation of the probability distribution of the variable $y = f(x)$, assuming that the random variable x has a probability distribution $g(x)$. However, this calculation is more general, and does not require the function $g(x)$ to be normalized to 1, or even to be normalizable. To compute an integral such as $I(a)$, the following transformation rule is used,

$$\delta(f(x) - a) = \sum_{i=1}^{n(a)} \frac{1}{|f'(x_i(a))|} \delta(x - x_i(a)) \tag{A.3}$$

where $x_1(a), \dots, x_{n(a)}(a)$ are the solutions of the equation $f(x) = a$ over the integration interval (x_{\min}, x_{\max}) . One thus ends up with the following expression for $I(a)$

$$I(a) = \int_{x_{\min}}^{x_{\max}} dx \sum_{i=1}^{n(a)} \frac{g(x)}{|f'(x_i(a))|} \delta(x - x_i(a)), \quad (\text{A.4})$$

leading after integration of the delta distributions to

$$I(a) = \sum_{i=1}^{n(a)} \frac{g(x_i(a))}{|f'(x_i(a))|}. \quad (\text{A.5})$$

Appendix B

Numerical Simulations of Markovian Stochastic Processes

In this appendix, we briefly describe some elementary methods to simulate Markovian stochastic processes. We first describe the easiest case of discrete time processes, and then move on to continuous time processes.

B.1 Discrete Time Processes

A discrete time Markovian stochastic process (also called Markov chain) is characterized by the list of transition probabilities $T(C'|C)$. We assume here that the process involves a finite number M of discrete configurations, which is often the case in practice. Configurations can thus be labelled as (C_1, \dots, C_M) . To simulate the stochastic dynamics, one needs to know how to choose a new configuration C' among (C_1, \dots, C_M) starting from an arbitrary configuration C . The new configuration C' has to be chosen randomly with a probability $T(C'|C)$. This can be done in practice in the following way. For a given configuration C , let us define the variables a_i

$$a_i = \sum_{j=1}^i T(C_j|C), \quad i = 1, \dots, M. \quad (\text{B.1})$$

One thus has by definition $a_M = 1$. It is also convenient to define $a_0 = 0$. We have for all $i = 1, \dots, M$ that $a_i - a_{i-1} = T(C_i|C)$. Drawing a random number u uniformly distributed over the interval $(0, 1]$, the probability that this random number falls between a_{i-1} and a_i is precisely $T(C_i|C)$, the length of the interval. Hence one simply has to determine i such that $a_{i-1} < u \leq a_i$, and to pick up the corresponding configuration C_i . In this way, the configuration C_i is indeed selected with a probability $T(C_i|C)$.

An efficient procedure to find the value i such that $a_{i-1} < u \leq a_i$ is to use a dichotomic algorithm. One starts from $j = E(M/2)$, where $E(x)$ is the integer part of x , and tests if $a_j < u$ or $a_j \geq u$. If $a_j < u$, the correct i satisfies $j + 1 \leq$

$i \leq M$, and one takes as a new trial value j' the middle of the interval, namely $j' = E((M + j + 1)/2)$. On the contrary, if $a_j \geq u$, the correct i satisfies $1 \leq i \leq j$, and the new trial value is $j' = E((j + 1)/2)$. By iteration, one rapidly converges to the value i satisfying $a_{i-1} < u \leq a_i$.

B.2 Continuous Time Processes

Continuous time Markovian processes are characterized by transition rates $W(C'|C)$, with $C' \neq C$. We assume here again that the process involves a finite number M of discrete configurations. Starting from a given configuration C_j , the questions are: (i) what is the probability to select the configuration C_i ($i \neq j$)? (ii) what is the time lag τ until the jump to the new configuration C_i ? The answer to point (i) is quite natural: configurations are selected with a probability proportional to the transition rates, meaning that the probability to choose configuration C_i starting from configuration C_j is

$$P(C_i|C_j) = \frac{W(C_i|C_j)}{\sum_{k(k \neq j)} W(C_k|C_j)}. \quad (\text{B.2})$$

Concerning point (ii), the time lag τ is a random variable following an exponential distribution

$$p(\tau) = \lambda_j e^{-\lambda_j \tau} \quad (\text{B.3})$$

where λ_j is the total 'activity'

$$\lambda_j = \sum_{i(i \neq j)} W(C_i|C_j). \quad (\text{B.4})$$

Hence to simulate the dynamics of a continuous time Markovian stochastic process, one has to draw a random number τ according to the exponential distribution (B.3), and to select a new configuration i with the probability $P(C_i|C_j)$ given in Eq. (B.2). The procedure to select the configuration is thus very similar to the one used in discrete time processes. The way to draw a random variable from an exponential distribution is explained in Appendix C. The algorithm to simulate continuous time Markovian stochastic processes is sometimes called the Gillespie algorithm.

Appendix C

Drawing Random Variables with Prescribed Distributions

Standard random number generators provide independent and identically distributed (pseudo-)random variables with a uniform distribution over the interval $(0, 1)$ —whether the boundaries 0 and 1 are included or not in the interval has to be checked case by case for each generator. The question encountered in practical simulations of stochastic processes is to be able to generate a random variable x with an arbitrary prescribed probability distribution $p(x)$, based on the uniform random number generator at hand. We describe below two methods enabling one to do so. More details can be found for instance in the standard textbook *Numerical Recipes* [1].

C.1 Method Based on a Change of Variable

The simplest method is based on a change of variable. For simplicity, we assume that the variable x is defined over an interval (a, b) , where $-\infty \leq a < b \leq +\infty$. Let us define the variable

$$u = F(x) \quad (a < x < b) \quad (\text{C.1})$$

with

$$F(x) \equiv \int_a^x p(x') dx' \quad (\text{C.2})$$

the cumulative distribution function of x . The probability distribution of u is denoted as $P(u)$, and is defined over the interval $(0, 1)$. The standard relation $P(u)|du| = p(x)|dx|$ connecting the distributions of u and x can be rewritten as

$$P(u) = \frac{p(x)}{|du/dx|}. \quad (\text{C.3})$$

From Eq. (C.1), we get $du/dx = p(x)$, so that we end up with $P(u) = 1$. Hence Eq. (C.1) connects a uniformly distributed variable to the desired variable x , and one

can simply generate x by drawing a uniform random number u and computing

$$x = F^{-1}(u) \quad (\text{C.4})$$

where F^{-1} is the reciprocal function of F . In practice, this method is useful only when an analytical expression of F^{-1} is available, which already covers a number of usual cases of interest, like exponential or power law distributions. For instance, an exponential distribution

$$p(x) = \lambda e^{-\lambda x} \quad (x > 0) \quad (\text{C.5})$$

with $\lambda > 0$ can be simulated using the change of variable

$$x = -\frac{1}{\lambda} \ln(1 - u). \quad (\text{C.6})$$

Since u and $(1 - u)$ have the same uniform distribution, one can in principle replace $(1 - u)$ by u in the r.h.s. of Eq. (C.6). One however needs to pay attention to the fact that the argument of the logarithm has to be non-zero, which guides the choice between u and $(1 - u)$, depending on whether 0 or 1 is excluded by the random number generator. Similarly, a power-law distribution

$$p(x) = \frac{\alpha x_0^\alpha}{x^{1+\alpha}} \quad (x > x_0) \quad (\text{C.7})$$

with $\alpha > 0$, can be simulated using

$$x = x_0 (1 - u)^{-1/\alpha}. \quad (\text{C.8})$$

Here again, the same comment about the choice of u or $(1 - u)$ applies. Many other examples where this method is applicable can be found.

When no analytical expression of the reciprocal function F^{-1} is available, one could think of using a numerical estimate of this function. There are however other more convenient methods that can be used in this case, as the rejection method described below.

Before describing this generic method, let us mention a generalization of the change of variable method, which as an important application allows for the simulation of a Gaussian distribution. Instead of making a change of variable on single variables, one can consider couples of random variables: $(x_1, x_2) = \mathcal{F}(u_1, u_2)$, where u_1 and u_2 are two independent uniform random numbers. It can be shown [1] that the following choice

$$\begin{aligned} x_1 &= \sqrt{-2 \ln u_1} \cos(2\pi u_2), \\ x_2 &= \sqrt{-2 \ln u_1} \sin(2\pi u_2), \end{aligned} \quad (\text{C.9})$$

leads to a pair of independent Gaussian random variables x_1 and x_2 , each with distribution

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (\text{C.10})$$

In practice, one often needs a single Gaussian variable at a time, and uses only one of the variables (x_1, x_2) . A Gaussian variable y of mean m and variance σ can be obtained by the simple rescaling $y = m + \sigma x$, where x satisfies the distribution (C.10).

C.2 Rejection Method

An alternative method, which is applicable to any distribution, is the rejection method that we now describe. Starting from an arbitrary target distribution $p(x)$ defined over an interval (a, b) (where a and/or b may be infinite), one first needs to find an auxiliary positive function $G(x)$ satisfying the three following conditions: (i) for all x such that $a < x < b$, $G(x) \geq p(x)$; (ii) $\int_a^b G(x) dx$ is finite; (iii) one is able to generate numerically a random variable x with distribution

$$\tilde{p}(x) = \frac{G(x)}{\int_a^b G(x') dx'} \quad (a < x < b), \quad (\text{C.11})$$

through another method, for instance using a change of variable. Then the rejection method consists in two steps. First, a random number x is generated according to the distribution $\tilde{p}(x)$. Second, x is accepted with probability $p(x)/G(x)$; this is done by drawing a uniform random number u over the interval $(0, 1)$, and accepting x if $u < p(x)/G(x)$. The geometrical interpretation of the rejection procedure is illustrated in Fig. C.1.

That the resulting variable x is distributed according to $p(x)$ can be shown using the following simple reasoning. Let us symbolically denote as A the event of drawing the variable x according to $\tilde{p}(x)$, and as B the event that x is subsequently accepted. We are interested in the conditional probability $P(A|B)$, that is, the probability distribution of the accepted variable. One has the standard relation

$$P(A|B) = \frac{P(A \cup B)}{P(B)}. \quad (\text{C.12})$$

The joint probability $P(A \cup B)$ is simply the product of the probability $\tilde{p}(x)$ and the acceptance probability $p(x)/G(x)$, yielding from Eq. (C.11)

$$P(A \cup B) = \frac{p(x)}{\int_a^b G(x') dx'}. \quad (\text{C.13})$$

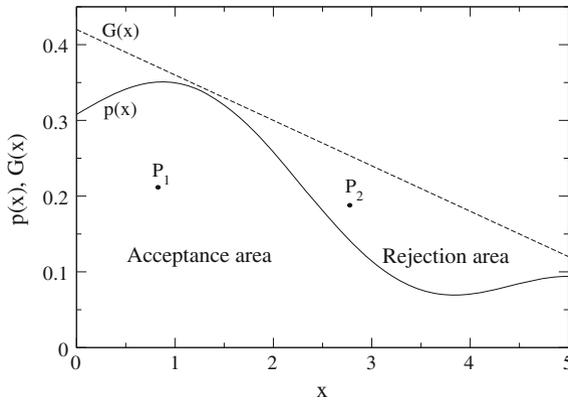


Fig. C.1 Illustration of the rejection method, aiming at drawing a random variable according to the normalized probability distribution $p(x)$ (full line). The function $G(x)$ (dashed line) is a simple upper bound of $p(x)$ (here, simply a linear function). A point P is randomly drawn, with uniform probability, in the area between the horizontal axis and the function $G(x)$. If P is below the curve defining the distribution p , its abscissa x is accepted (point P_1); it is otherwise rejected (point P_2). The random variable x constructed in this way has probability density $p(x)$ —see text

Then, $P(B)$ is obtained by summing $P(A \cup B)$ over all events A , yielding

$$P(B) = \int_a^b dx \frac{p(x)}{\int_a^b G(x') dx'} = \frac{1}{\int_a^b G(x') dx'}. \quad (\text{C.14})$$

Combining Eqs. (C.12)–(C.14) eventually leads to $P(A|B) = p(x)$.

From a theoretical viewpoint, any function satisfying conditions (i), (ii) and (iii) is appropriate. Considering the efficiency of the numerical computation, it is however useful to minimize the rejection rate, equal from Eq. (C.14) to

$$r = 1 - \frac{1}{\int_a^b G(x) dx}. \quad (\text{C.15})$$

Hence the choice of the function $G(x)$ should also try to minimize $\int_a^b G(x) dx$, to make it relatively close to 1 if possible. Note that $G(x)$ does not need to be a close upper approximation of $p(x)$ everywhere, only the integral of $G(x)$ matters.

Reference

1. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes, The Art of Scientific Computing*, 3rd edn. (Cambridge University Press, Cambridge, 2007)