Jan Naudts

Generalised Thermostatistics

Springer

Generalised Thermostatistics

Jan Naudts

Generalised Thermostatistics

Springer

Jan Naudts
University of Antwerp
Physics Department
Universiteitsplein 1
2610 Antwerp
Belgium
jan.naudts@ua.ac.be

ISBN 978-0-85729-354-1 e-ISBN 978-0-85729-355-8
DOI 10.1007/978-0-85729-355-8
Springer London Dordrecht Heidelberg New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2011922042

Mathematics Subject Classification (2010): 82B05, 82B10, 82B30

Cover design: deblik

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is in the first place a monograph, for a large part based on my own research. The same notations are used throughout the text and a strict unity of subject is attempted. On the other hand, care is taken that the book can be used as a textbook for graduate students and as a reference text for researchers.

The point of view adopted here is that statistical physics is statistics applied to physics. The essence of statistical physics is the statistical analysis of models which approximate the physical reality. These models consist of a probability distribution, or its quantum mechanical analogue, used to calculate statistical averages. The probability distribution depends on a small number of parameters which can be estimated from experimental data.

The roots of statistical physics lie in thermodynamics, a nineteenth century science predating statistical physics. One of the goals of statistical physics is to explain thermodynamical concepts in terms of a microscopic theory. But thermodynamics is so generally valid that its relations still hold in a much wider context than that of the traditional theory of statistical mechanics. Therefore, some notions of thermodynamics appear as a skeleton throughout the book.

Recent years, considerable efforts have been made to extend the statistical physics formalism beyond the limits set out by Gibbs [5] in his book, published in 1902. Traditional statistical physics focuses on systems with many degrees of freedoms. The formalism becomes exact in the *thermodynamic limit*, this is, the limit of infinitely many degrees of freedom. One motivation to go beyond the standard formalism is the current interest in relatively small systems. Many new insights originate from *Tsallis' non-extensive thermostatistics*, a domain of research that developed during the past twenty years. Let me mention in particular the notions of *deformed exponential and logarithmic functions*, and of *escort probability distributions*, notions that play an important role in Part II of the book.

The title of this book refers to the well-known book by Callen, "Thermodynamics and an Introduction to Thermostatistics" [3], which is often cited in that part of literature which is concerned with non-extensive thermostatistics.

The emphasis in the present work lies on the development of the formalism. For applications of non-extensive statistical physics the reader is referred to the book by Constantino Tsallis [7], to proceedings of conferences [4, 1], and to some review papers [6, 2]. A number of topics, playing a central role in traditional statistical physics, are not treated in the present text. Let me mention the *equivalence of ensembles*, the *thermodynamic limit*, the *central limit theorem*, *large deviation theory*. Time-dependent phenomena are not discussed. The main reason for the latter limitation is that *non-equilibrium statistical physics* is an active area of research with only recent understanding of some, not all, of its fundamentals.

This book does not intend to review the research literature on non-extensive thermostatistics. It rather situates this subject in a broader context and aims at consolidating its results. The short notes at the end of each Chapter try to indicate some aspects of the historical developments but fall short of giving proper credit to all researchers active in the field.

I am grateful to all colleagues who helped me to improve the contents of this book. I am especially indebted to Marek Chachor, who introduced me to non-extensive statistical physics.

Antwerpen, September 2010 *Jan Naudts*

References

1. Abe, S., Herrmann, H., Quarati, P., Rapisarda, A., Tsallis, C. (eds.): Complexity, metastability and nonextensivity, *AIP Conference Proceedings*, vol. 965. American Institute of Physics (2007) vi
2. Beck, C.: Generalised information and entropy measures in physics. Contemporary Physics **50**, 495–510 (2009) vi
3. Callen, H.: Thermodynamics and an Introduction to Thermostatistics, 2nd edn. Wiley, New York (1985) v, 23, 37
4. Gell-Mann, M., Tsallis, C. (eds.): Nonextensive Entropy. Oxford University Press, Oxford (2004) vi
5. Gibbs, J.W.: Elementary principles in statistical mechanics. Reprint. Dover, New York (1960) v, 5, 56, 66
6. Tsallis, C.: Nonextensive statistical mechanics: construction and physical interpretation. In: M. Gell-Mann, C. Tsallis (eds.) Nonextensive Entropy, pp. 1–53. Oxford University Press, Oxford (2004) vi
7. Tsallis, C.: Introduction to nonextensive statistical mechanics. Springer (2009) vi

Contents

Part I Parameter Estimation

1 Probability Distributions 3
1.1 The Maxwell Distribution 3
1.2 Probability Distributions in Phase Space 4
1.3 The Boltzmann-Gibbs Distribution 5
1.4 A Gas of Particles in the Canonical Ensemble 6
1.5 Additional Conserved Quantities 8
1.6 The Grand-Canonical Ensemble 10
1.7 Quantum Statistics 12
1.8 The von Neumann Density Operator 14
1.9 Fermi-Dirac and Bose-Einstein Distributions 16
Problems 18
Notes 19
Objectives 20
References 20

2 Statistical Models 21
2.1 Parameter Estimation 21
2.2 Definition of a Statistical Model 23
2.3 The Exponential Family 25
2.4 Curved Exponential Families 27
2.5 Example: The Ising Model in d=1 29
2.6 The Density of States 31
2.7 The Quantum Exponential Family 32
Problems 34
Notes 36
Objectives 36
References 36

3 Thermodynamic Equilibrium 37
3.1 Thermodynamic Configuration Space 37
3.2 Maximum Entropy Principle 38
3.3 BGS Entropy Functional 39
3.4 Applying the Method of Lagrange 41
3.5 Thermodynamic Entropy 42
3.6 Relative Entropy 44
3.7 Thermodynamic Stability 45
3.8 Entropy of Probability Densities 47
3.9 Quantum Entropies 47
Problems 49
Notes 50
Objectives 51
References 51

4 The Microcanonical Ensemble 53
4.1 Introduction 53
4.2 The Ergodic Theorem 54
4.3 Example: The Harmonic Oscillator 55
4.4 Definition 56
4.5 Microcanonical Instabilities 57
4.6 The Quantum Microcanonical Ensemble 62
4.7 The Coherent State Ensemble 64
Problems 65
Notes 66
Objectives 67
References 67

5 Hyperensembles 69
5.1 Introduction 69
5.2 The Canonical Ensemble 70
5.3 Superstatistics 71
5.4 The Hyperensemble 74
5.5 Properties of the Hyperensemble 75
Notes 77
Objectives 78
References 78

6 The Mean Field Approximation 79
6.1 The Ideal Paramagnet 79
6.2 The Mean Field Equation 81
6.3 Phase Transitions 82
6.4 A Mean Field Phase Transition 83
6.5 A Hyperensemble of Product States 86
6.6 Generalised Mean Field Theories 86

6.7 The Quantum Case ... 89
Notes ... 91
Objectives ... 91
References ... 91

Part II Deformed Exponential Families

7 q-Deformed Distributions ... 95
7.1 q-Deformed Exponential and Logarithmic Functions ... 95
7.2 Dual Definitions ... 97
7.3 The q-Exponential Family ... 98
7.4 Escort Probabilities ... 100
7.5 q-to-1/q Duality and Escort Families ... 101
7.6 Dual Identities ... 102
7.7 The q-Gaussian Distribution ... 103
7.8 Configurational Probability Distribution ... 106
7.9 Average Kinetic Energy ... 108
7.10 The Quantum Family ... 109
Problems ... 110
Notes ... 112
Objectives ... 113
References ... 113

8 Tsallis' Thermostatistics ... 115
8.1 The Tsallis Entropy Functional ... 115
8.2 A Historical Reflection ... 117
8.3 Maximising the Tsallis Entropy Functional ... 118
8.4 Thermodynamic Properties ... 119
8.5 Example: The Two-Level Atom ... 121
8.6 Relative Entropy of the Csiszár Type ... 123
8.7 Relative Entropy of the Bregman Type ... 124
8.8 Quantum Expressions ... 127
8.9 More General Entropies ... 127
Problems ... 128
Notes ... 129
Objectives ... 129
References ... 130

9 Changes of Scale ... 131
9.1 Kolmogorov-Nagumo Averages ... 131
9.2 Rényi's Alpha Entropies ... 132
9.3 Rényi or Tsallis? ... 133
9.4 Configurational Temperature ... 135
9.5 Fractal Dimensions ... 137
9.6 Microcanonical Description ... 140

9.7 Sharma-Mittal Entropy Functional 143
9.8 Zipf's Law 145
Problems 146
Notes 147
Objectives 147
References 148

10 General Deformations 149
10.1 Deformed Exponential and Logarithmic Functions 149
10.2 Dual Definitions 150
10.3 Deduced Logarithms 151
10.4 The Phi-Exponential Family 153
10.5 Escort Probabilities 153
10.6 Method to Test Phi-Exponentiality 155
10.7 Example: The Site Percolation Problem 157
10.8 Generalised Quantum Statistics 160
Problems 161
Notes 162
Objectives 162
References 162

11 General Entropies 165
11.1 The Entropy Functional 165
11.2 Relative Entropies 166
11.3 The Variational Principle 167
11.4 Complexity 169
11.5 Application to q-Deformed Distributions 170
11.6 Deformed Fisher Information 171
11.7 Quantum Entropies 173
11.8 Quantum Stability 174
Problems 176
Notes 177
Objectives 177
References 178

Solutions to the Problems 179

Index 199

Part I
Parameter Estimation

The point of view developed in this part of the book is that statistical mechanics is statistics applied to classical mechanics or quantum mechanics. The essential concept is that of the exponential family. It gives a mathematically concise formulation of the Boltzmann-Gibbs distribution.

One of the goals of statistical physics is to give a microscopic basis to thermodynamics. Some notions of thermodynamics are introduced in the Section 3. Because of the essential role of the micro-canonical ensemble within statistical mechanics the emphasis lies on the entropy $S(U)$ as a function of the energy U, and its Legendre transform, which is Massieu's function. This function replaces the free energy $F(T)$ which is a function of the temperature T and which is the Legendre transform of the energy $U(S)$ as a function of the entropy S.

Chapter 1
Probability Distributions in Statistical Physics

1.1 The Maxwell Distribution

A cubic meter of air contains of the order of 2.5×10^{25} molecules (about 40 moles, one mole contains per definition $N_A \simeq 6 \times 10^{23}$ molecules — this is Avogadro's number). Because of this tremendously large number, it is a good idea to do statistics on the speed v of individual molecules. For a gas of identical molecules, each with mass m, one finds in very good approximation that the probability density is of the form

$$f(v) = \left(\frac{\beta m}{2\pi}\right)^{3/2} 4\pi v^2 e^{-\frac{1}{2}\beta m v^2}. \tag{1.1}$$

See Figure 1.1. In statistical physics, β is a shorthand notation for the inverse of the temperature T, multiplied with *Boltzmann's constant* $k_B \simeq 1.38066 \times 10^{-23}$ J/K, to convert it from degrees Kelvin to energy units (Joule)

$$\beta = \frac{1}{k_B T}. \tag{1.2}$$

The density function (1.1) is known as the *Maxwell distribution*.

Note that (1.1) is properly normalised

$$\int_0^{+\infty} \mathrm{d}v\, f(v) = 1. \tag{1.3}$$

The first moment of $f(v)$ is the average speed $\langle v \rangle$. Its value is of the same order of magnitude as the speed of sound, which for air is about 330 m/s (about 1200 km per hour). The calculation of $\langle v \rangle$ requires the evaluation of an integral. The second moment is easier to calculate. Indeed, by taking the derivative of (1.3) with respect to β, there follows

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_1,

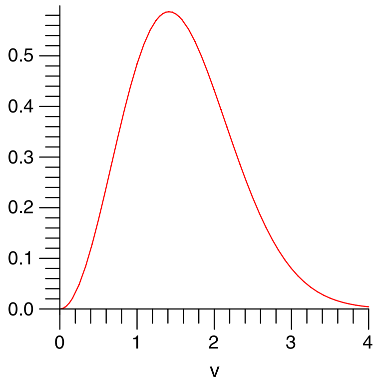

Fig. 1.1 The Maxwell distribution with $\beta m = 1$

$$0 = \frac{\partial}{\partial \beta} \int_0^{+\infty} \mathrm{d}v\, f(v) \\ = \frac{3}{2\beta} - \frac{1}{2} m\langle v^2 \rangle. \tag{1.4}$$

This can be written as

$$\frac{1}{2} m\langle v^2 \rangle = \frac{3}{2} k_B T. \tag{1.5}$$

Now, $\frac{1}{2}mv^2$ is in classical mechanics the kinetic energy of a particle with mass m and speed v. Hence, (1.5) says that the average kinetic energy per particle is proportional to temperature T. The result (1.4) is known as the *equipartition law*, and is related to the law of Dulong and Petit.

Expression (1.5) is quite often used as the definition of temperature T, especially in the context of experiments. However, this cannot be the fundamental definition since we know that quantum mechanics brings corrections to (1.5).

1.2 Probability Distributions in Phase Space

The Maxwell distribution is used for classical gases but considers only the speed of one molecule of the gas. A powerful generalisation is obtained by considering at once the positions and momenta of all molecules of the gas. This involves probability distributions in phase space.

A *probability distribution in the phase space* Γ, is a positive function $f(q, p)$ depending on the coordinates $q_1, \cdots, q_N$ and momenta $p_1, \cdots, p_N$ of all N particles of the system. Its normalisation

$$\frac{1}{N!h^{3N}}\int_{\Gamma}\mathrm{d}q\mathrm{d}p\, f(q,p) = 1 \tag{1.6}$$

involves a constant h with the same dimensions as the product of a position and a momentum variable. In this way the density function becomes dimensionless. One often chooses h equal to Planck's constant $h \simeq 6.625 \times 10^{-34}$ Joule Sec, although within classical mechanics there is no fundamental reason to do so. Alternatively, $h = 1$ is used. This is equivalent with measuring momenta in units of inverse length. For use later on, in the section on the grand-canonical ensemble, a factor $1/N!$ has been added in the normalisation. It reflects that all particles are identical and that each of the $N!$ permutations of particles describes the same system.

Given any such probability distribution one can calculate averages of other functions of phase space $A(q,p)$ using the obvious formula

$$\langle A\rangle = \frac{1}{N!h^{3N}}\int_{\Gamma}\mathrm{d}q\mathrm{d}p\, f(q,p)A(q,p). \tag{1.7}$$

1.3 The Boltzmann-Gibbs Distribution

In the second half of the nineteenth century, Boltzmann has generalised the Maxwell distribution by introducing a probability distribution in phase space. This generalised distribution is now called the Boltzmann-Gibbs distribution. It plays a central role in the foundations of statistical mechanics. An account of the latter is found in the book by J. Willard Gibbs [1], published in 1902.

The specific choice of probability distribution in phase space that carries the name of Boltzmann and Gibbs is

$$f(q,p) = \frac{1}{Z(\beta)}e^{-\beta H(q,p)}. \tag{1.12}$$

The normalisation $Z(\beta)$ is called the *partition sum* and is given by

$$Z(\beta) = \frac{1}{N!h^{3N}}\int_{\Gamma}\mathrm{d}q\mathrm{d}p\, e^{-\beta H(q,p)}. \tag{1.13}$$

As before, β is the inverse temperature. The function $H(q,p)$ is Hamilton's function. Its value is the total energy of a system where the particles have positions $q_1,\ q_2,\ \cdots q_N$ and associated momenta $p_1,\ p_2,\ \cdots p_N$. A system described by the Boltzmann-Gibbs distribution (1.12) is said to belong to the *canonical ensemble*.

Note that averages calculated using the Boltzmann-Gibbs distribution do not depend on time. See the Box 1.1.

The state of a system of classical mechanics is described by a number of coordinates $q_1, q_2, \cdots q_N$ and associated momenta $p_1, p_2, \cdots p_N$. Together they determine a point (q, p) in the phase space $\Gamma \subset \mathbf{R}^{6N}$. The time evolution of the system is determined by *Hamilton's equations of motion*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} q_{j\alpha} &= \frac{\partial H}{\partial p_{j\alpha}} \\ \frac{\mathrm{d}}{\mathrm{d}t} p_{j\alpha} &= -\frac{\partial H}{\partial q_{j\alpha}}.\end{aligned} \tag{1.8}$$

The function $H(q, p)$ is the Hamiltonian. Its value is the energy at the given point (q, p) of phase space. The index α runs from 1 to 3, or from x to z.

Introduce now time-dependent functions by posing that

$$A_t(q(0), p(0)) \equiv A(q(t), p(t)) \tag{1.9}$$

(usually the time-dependence of q and p is implicit — here it is written explicitly to make clear that the time-dependence is shifted from the coordinates to the function). Then the Boltzmann-Gibbs expectation of the function A at time t is given by

$$\langle A_t \rangle = \frac{1}{N! h^{3N}} \int_\Gamma \mathrm{d}q\mathrm{d}p \, \frac{1}{Z(\beta)} e^{-\beta H(q,p)} A_t(q, p). \tag{1.10}$$

Because the Hamiltonian $H(q, p)$ is a conserved quantity one can write as well

$$\langle A_t \rangle = \frac{1}{N! h^{3N}} \int_\Gamma \mathrm{d}q\mathrm{d}p \, \frac{1}{Z(\beta)} e^{-\beta H_t(q,p)} A_t(q, p). \tag{1.11}$$

By the *theorem of Liouville*, the integral over phase space of the function $e^{-\beta H_t(q,p)} A_t(q, p)$ does not depend on time. One therefore concludes that $\langle A_t \rangle = \langle A \rangle$ does not depend on time.

Box 1.1 Classical mechanics

1.4 A gas of particles in the canonical ensemble

A classical gas composed of N identical particles, all with identical mass m, is described by a Hamiltonian of the form

$$H = \frac{1}{2m} \sum_{j=1}^{N} \sum_{\alpha=x,y,z} p_{j\alpha}^2 + \mathcal{V}(q). \tag{1.14}$$

The momentum and position of the particle labelled j each have three components $p_{j,x}$, $p_{j,y}$,$p_{j,z}$, respectively $q_{j,x}$, $q_{j,y}$, $q_{j,z}$. The first contribution is the kinetic energy of the particles. The potential energy $\mathcal{V}(q)$ is a function of positions only. It describes the interaction between the particles. The partition sum of this system is given by

$$Z(\beta) = \frac{1}{N!h^{3N}} \int_{\mathbf{R}^{3N}} \mathrm{d}p \int_{V^N} \mathrm{d}q \, \exp\left(-\frac{\beta}{2m} \sum_{j=1}^{N} \sum_{\alpha=x,y,z} p_{j\alpha}^2 - \beta\mathcal{V}(q)\right). \tag{1.15}$$

$V \subset \mathbf{R}^3$ is the region of space enclosing the particles. By abuse of notation V will also denote the volume of this region of space. The integrations over the momentum variables can be carried through. The result is

$$Z(\beta) = \frac{1}{N!h^{3N}} \left(2\pi m k_B T\right)^{3N/2} Z_{\text{conf}}(\beta), \tag{1.16}$$

with

$$Z_{\text{conf}}(\beta) = \int_{V^N} \mathrm{d}q \, \exp\left(-\beta\mathcal{V}(q)\right). \tag{1.17}$$

The function $Z_{\text{conf}}(\beta)$ is called the *configurational partition sum*. For an arbitrary function $A(q)$ depending only on the positions of the particles is

$$\langle A \rangle = \frac{1}{Z_{\text{conf}}(\beta)} \int_{V^N} \mathrm{d}q \, \exp\left(-\beta\mathcal{V}(q)\right) A(q). \tag{1.18}$$

It is now straightforward to show that the Maxwell distribution is a special case of Boltzmann-Gibbs. See the Box (1.2).

If the interaction between the particles is absent ($\mathcal{V}(q) \equiv 0$) then the gas is said to be *ideal*. Note that in that case one has $Z_{\text{conf}}(\beta) = V^N$ so that

$$Z(\beta) = \frac{1}{N!h^{3N}} \left(2\pi m k_B T\right)^{3N/2} V^N. \tag{1.22}$$

The trick (1.4), which, starting from the normalisation of the Maxwell distribution, leads to the *equipartition law*, can now be repeated. It requires two steps. At one hand, the explicit result (1.22) can be used to obtain

$$\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta) = -\frac{3N}{2} k_B T. \tag{1.23}$$

On the other hand, a formal calculation starting from (1.13) gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta) &= \frac{1}{Z(\beta} \frac{\mathrm{d}}{\mathrm{d}\beta} \frac{1}{N!h^{3N}} \int_{\Gamma} \mathrm{d}q\mathrm{d}p \, e^{-\beta H(q,p)} \\
&= -\frac{1}{N!h^{3N}} \int_{\Gamma} \mathrm{d}q\mathrm{d}p \, e^{-\beta H(q,p)} H(q,p) \\
&= -\langle H \rangle.
\end{aligned} \tag{1.24}$$

For an arbitrary function $A(p_j)$ depending only on the momenta of the j-th particle is

$$\langle A\rangle = \frac{\int_{\mathbf{R}^3} \mathrm{d}p_j \, \exp\left(-\frac{\beta}{2m}\sum_{\alpha=x,y,z} p_{j\alpha}^2\right) A(p_j)}{\int_{\mathbf{R}^3} \mathrm{d}p_j \, \exp\left(-\frac{\beta}{2m}\sum_{\alpha=x,y,z} p_{j\alpha}^2\right)}. \tag{1.19}$$

Note that the speed v of this particle satisfies $\sum_{\alpha=x,y,z} p_{j\alpha}^2 = m^2v^2$. Hence, if the function A depends only on v, then one obtains

$$\langle A\rangle = \frac{\int_{\mathbf{R}^3} \mathrm{d}p_j \, \exp\left(-\frac{1}{2}\beta m v^2\right) A(v)}{\int_{\mathbf{R}^3} \mathrm{d}p_j \, \exp\left(-\frac{1}{2}\beta m v^2\right)}. \tag{1.20}$$

Replacing the integration over Cartesian coordinates $p_{jx}.p_{jy}, p_{jz}$ by an integration in spherical coordinates gives

$$\langle A\rangle = \frac{\int_0^{+\infty} v^2 \mathrm{d}v \, \exp\left(-\frac{1}{2}\beta m v^2\right) A(v)}{\int_0^{+\infty} v^2 \mathrm{d}v \, \exp\left(-\frac{1}{2}\beta m v^2\right)}. \tag{1.21}$$

The latter expression implies that the speed v obeys a Maxwell distribution.

Box 1.2 Derivation of the Maxwell distribution

This shows that the average energy of an ideal gas in the canonical ensemble equals $(3N/2)k_BT$. This result is in agreement with (1.5) because for an ideal gas, there is no potential energy. Hence, the total energy equals the kinetic energy.

1.5 Additional Conserved Quantities

The Boltzmann-Gibbs distribution (1.12) depends on the single parameter β. A straightforward way to further generalise this distribution is by introducing more than one parameter. An argument for doing so in statistical physics is the occurrence of additional conserved quantities in the study of certain models. Indeed, the Hamiltonian $H(q,p)$ that appears in (1.12) does not depend on time. For that reason it is an important quantity characterising the macroscopic state of the system. If an additional quantity exists which is used to characterise the macroscopic state of the system then it is obvious to include it somehow in the probability distribution (1.12).

Consider as an example a cylinder filled with gas, in the presence of a uniform gravitational force. See Figure 1.2. The Hamiltonian is of the form (1.14), with the potential energy $\mathcal{V}(q)$ given by

$$\mathcal{V}(q) = \mathcal{V}_0(q) + gm \sum_{n=1}^{N} q_{nz}. \tag{1.25}$$

The first term describes the interactions between the molecules of the gas, the last term describes a uniform acceleration of all particles in the negative z-direction. The quantity $Q_z(q) = \frac{1}{N}\sum_{n=1}^{N} q_{nz}$ is the average height of a molecule and is a constant of motion. One can of course continue to describe this gas using the Boltzmann-Gibbs distribution (1.12). However, the alternative is to rewrite (1.12) as

$$f(q,p) = \frac{1}{Z(\beta,g)} e^{-\beta H_0(q,p) - \beta g m N Q_z(q)}, \tag{1.26}$$

with $H_0(q,p)$ the value of $H(q,p)$ when $g = 0$. The model now depends on two parameters β and g. One advantage of the two-parameter model is that the trick, used to derive (1.5) and (1.24), can now be applied for both parameters. Indeed, from the normalisation (1.6) now follows

$$\begin{aligned} 0 &= \frac{\partial}{\partial\beta} \int_\Gamma \mathrm{d}q\mathrm{d}p\, f(q,p) \\ &= -\frac{1}{Z(\beta,g)} \frac{\partial}{\partial\beta} Z(\beta,g) - \int_\Gamma \mathrm{d}q\mathrm{d}p\, f(q,p) H(q,p), \end{aligned} \tag{1.27}$$

and similarly

$$\begin{aligned} 0 &= \frac{\partial}{\partial g} \int_\Gamma \mathrm{d}q\mathrm{d}p\, f(q,p) \\ &= -\frac{1}{Z(\beta,g)} \frac{\partial}{\partial g} Z(\beta,g) - \int_\Gamma \mathrm{d}q\mathrm{d}p\, f(q,p) \beta m N Q_z(q). \end{aligned} \tag{1.28}$$

These expressions can be written as

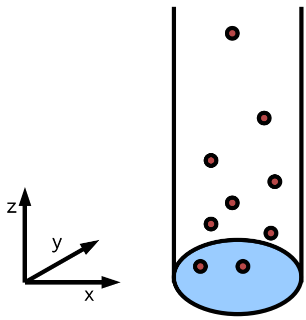

Fig. 1.2 A gas in a uniform field

$$\frac{\partial}{\partial \beta} \ln Z(\beta, g) = -\langle H \rangle,$$
$$\frac{\partial}{\partial g} \ln Z(\beta, g) = -\beta m N \langle Q_z \rangle. \tag{1.29}$$

Hence, by calculating the partition sum for the two-parameter model one gets easy access to the average values of the two quantities H and Q_z. From

$$\ln Z(\beta, g) = -\frac{5}{2} N \ln \beta - N \ln g + \text{constant} \tag{1.30}$$

follows $\langle H \rangle = \frac{5}{2} N k_B T$ and $\langle Q_z \rangle = k_B T / mg$.

It may seem weird to calculate averages of conserved quantities since by definition they have a constant value and their value does not fluctuate as a function of time. However, quite often the exact value of a conserved quantity is not known so that the uncertainty about its value can be treated in a statistical way. Alternatively, one may say that a physical system is never completely isolated. As a consequence of the interactions with the environment, called the *heat bath* , the quantity, which is conserved in the isolated system, has a fluctuating value. In thermodynamics, these variables are called *extensive*, because usually their value is proportional to the size of the system. The average value of the extensive variable is then determined by a *control parameter*, also called an *intensive variable*. In the previous example, the inverse temperature β controls the expected value of the energy, the strength of the uniform field g controls the average height of the molecules.

1.6 The Grand-Canonical Ensemble

One special case of a situation with an additional conserved quantity concerns the number of particles N. It is special because every time one changes N one switches to another phase space Γ, and to another Hamiltonian $H(q, p)$, defined for points (q, p) in Γ. To make this point explicit, we add an index N to the Hamiltonian and to the phase space symbol, so $H_N(q, p)$ and Γ_N. This index N makes clear that the argument (q, p) belongs to Γ_N, and hence has $6N$ components.

The situation of interest is a system, for instance a gas, in which the number of particles N is not known precisely and is therefore treated statistically. The corresponding control parameter is the *chemical potential* μ. In daily life, this parameter is less known than the temperature. Instead one uses the pressure of a gas as the quantity controlling the number of particles. As we will see below, the pressure has a different role in the formalism of statistical physics.

Ideally, the number of gas particles in a container can vary between 0 and infinity. In the *grand-canonical ensemble*, which is the formalism that is now described, all these possible values of N are considered simultaneously. This goal is reached by considering an infinite row of density distributions $f_N(q,p)$, $N = 0, 1, 2, \cdots$, normalised in such a way that

$$1 = \sum_{N=0}^{\infty} \frac{1}{N!h^{3N}} \int_{\Gamma_N} \mathrm{d}q\mathrm{d}p\, f_N(q,p). \tag{1.31}$$

Note that the first contribution to the infinite sum equals f_0 and is the probability that the system contains zero particles. Using this row of density distributions one can calculate averages of quantities which depend on the number of particles N and, given N, on the point (q,p) in the phase space Γ_N, by the expression

$$\langle A \rangle = \sum_{N=0}^{\infty} \frac{1}{N!h^{3N}} \int_{\Gamma_N} \mathrm{d}q\mathrm{d}p\, f_N(q,p) A_N(q,p). \tag{1.32}$$

Examples of functions $A_N(q,p)$ are the energy $H_N(q,p)$ and the number of particles N (the latter does not depend on q and p).

The Boltzmann-Gibbs distribution in this case becomes, similarly to (1.26),

$$f_N(q,p) = \frac{1}{Z(\beta,\mu)} e^{-\beta(H_N(q,p)-\mu N)}. \tag{1.33}$$

The *grand-canonical partition sum* must be such that the average of a constant is the constant itself. Taking $A \equiv 1$ in (1.32) yields

$$\begin{aligned} Z(\beta,\mu) &= \sum_{N=0}^{\infty} e^{\beta\mu N} \frac{1}{N!h^{3N}} \int_{\Gamma_N} \mathrm{d}q\mathrm{d}p\, e^{-\beta H_N(q,p)} \\ &= \sum_{N=0}^{\infty} e^{\beta\mu N} Z_N(\beta), \end{aligned} \tag{1.34}$$

with $Z_N(\beta)$ the canonical partition sum for a system with N particles (By convention put $Z_0(\beta) = 1$).

Now one calculates

$$\begin{aligned} 0 &= \frac{\partial}{\partial\beta} \sum_{N=0}^{\infty} \frac{1}{N!h^{3N}} \int_{\Gamma_N} \mathrm{d}q\mathrm{d}p\, \frac{1}{Z(\beta,\mu)} e^{-\beta(H_N(q,p)-\mu N)} \\ &= -\frac{1}{Z(\beta,\mu)} \frac{\partial}{\partial\beta} Z(\beta,\mu) - \langle H - \mu N \rangle, \end{aligned} \tag{1.35}$$

and similarly

$$0 = \frac{\partial}{\partial\mu} \sum_{N=0}^{\infty} \frac{1}{N!h^{3N}} \int_{\Gamma_N} \mathrm{d}q\mathrm{d}p\, \frac{1}{Z(\beta,\mu)} e^{-\beta(H_N(q,p)-\mu N)}$$

$$= -\frac{1}{Z(\beta,\mu)}\frac{\partial}{\partial\mu}Z(\beta,\mu) + \beta\langle N\rangle. \tag{1.36}$$

Hence, one has the identities

$$\frac{\partial}{\partial\beta}\ln Z(\beta,\mu) = -\langle H\rangle + \mu\langle N\rangle, \tag{1.37}$$

$$\frac{\partial}{\partial\mu}\ln Z(\beta,\mu) = \beta\langle N\rangle. \tag{1.38}$$

In the case of an *ideal gas* one can calculate $Z(\beta,\mu)$ explicitly. One finds (see (1.22))

$$\begin{aligned}
\ln Z(\beta,\mu) &= \ln\sum_{N=0}^{\infty} e^{\beta\mu N} Z_N(\beta) \\
&= \ln\sum_{N=0}^{\infty} e^{\beta\mu N}\frac{1}{N!h^{3N}}\left(2\pi m k_B T\right)^{3N/2} V^N \\
&= \frac{1}{h^3}e^{\beta\mu}(2\pi m k_B T)^{3/2}V.
\end{aligned} \tag{1.39}$$

This implies, using the identities (1.37, 1.38),

$$\begin{aligned}
\langle H\rangle - \mu\langle N\rangle &= (\frac{3}{2}k_B T - \mu)\ln Z(\beta,\mu) \\
\beta\langle N\rangle &= \beta\ln Z(\beta,\mu).
\end{aligned} \tag{1.40}$$

Eliminating $\ln Z(\beta,\mu)$ gives the *equipartition law*

$$\langle H\rangle = \frac{3}{2}k_B T\langle N\rangle. \tag{1.41}$$

On the other hand, comparing $\langle N\rangle = \ln Z(\beta,\mu)$ with the *ideal gas law* $pV = \langle N\rangle k_B T$, where p is the *pressure* of the ideal gas, yields

$$pV = k_B T\ln Z(\beta,\mu). \tag{1.42}$$

The latter expression shows that the logarithm of the grand-canonical partition sum is proportional to the pressure of the gas.

1.7 Quantum Statistics

The observables of a quantum mechanical model are self-adjoint operators of a Hilbert space $\mathcal{H}$. Averages of observables are determined by a *density operator* ρ (often called a *density matrix*). This is a trace-class operator satisfying the two conditions of positivity and normalisation

In quantum mechanics the Hamiltonian H is a self-adjoint operator in a Hilbert space $\mathcal{H}$. It is the generator of time evolution, which means it determines unitary operators $U(t)$ by

$$U(t) = e^{-i\hbar^{-1}tH}. \tag{1.43}$$

The state of the system is described by a normalised element ψ of $\mathcal{H}$ and is called a *wave function*. Its time evolution is given by

$$\psi(t) = U(t)\psi. \tag{1.44}$$

The time evolution of an operator A in the Heisenberg picture is given by

$$A(t) = U(t)^* A U(t). \tag{1.45}$$

Using the density operator of von Neumann one has

$$\begin{aligned} \langle A(t) \rangle &= \operatorname{Tr} \rho A(t) \\ &= \operatorname{Tr} \rho U(t)^* A U(t) \\ &= \operatorname{Tr} U(t) \rho U(t)^* A \\ &= \operatorname{Tr} \rho U(t) U(t)^* A \\ &= \operatorname{Tr} \rho A = \langle A \rangle. \end{aligned} \tag{1.46}$$

One concludes that thermal averages calculated using the density operator of von Neumann do not depend on time. To obtain this result we used cyclic permutation under the trace, the fact that $U(t)$ and ρ commute, and $U(t)U(t)^* = \mathbf{I}$.

Box 1.3 Quantum mechanics

$$\rho \geq 0 \qquad \text{and} \qquad \operatorname{Tr} \rho = 1. \tag{1.47}$$

There exists always a basis of eigenfunctions ϕ_n of ρ: $\rho\phi_n = \lambda_n \phi_n$. In this basis, ρ is diagonal and can be written as

$$\rho = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & \cdots \\ 0 & \lambda_2 & \cdots & 0 & \cdots \\ & & \cdots & & \\ 0 & 0 & \cdots & \lambda_n & \cdots \\ \cdots & & & & \end{pmatrix}, \tag{1.48}$$

with $\lambda_n \geq 0$ and $\sum_n \lambda_n = 1$.

The average of the bounded operator A is then defined by

$$\langle A \rangle = \operatorname{Tr} \rho A. \tag{1.49}$$

This definition of average generalises the concept of statistical averages to the quantum context. It satisfies two basic axioms

i) The average of a constant is the constant itself. More precisely, if the operator A is a constant λ times the identity operator $\mathbf{I}$, then one has $\langle A \rangle = \langle \lambda \mathbf{I} \rangle = \operatorname{Tr} \rho \lambda \mathbf{I} = \lambda$ because the trace operation is linear and $\operatorname{Tr} \rho = 1$.
ii) The average of a positive quantity is non-negative. Indeed, If $A \geq 0$ then it has a square root $A^{1/2}$. Hence, using cyclic permutation under the trace, one obtains

$$\begin{aligned} \langle A \rangle = \operatorname{Tr} \rho A &= \operatorname{Tr} A^{1/2} \rho A^{1/2} \\ &= \operatorname{Tr} (A^{1/2} \rho^{1/2})(A^{1/2} \rho^{1/2})^* \\ &\geq 0. \end{aligned} \tag{1.50}$$

In quantum statistics, the notion of conditional probability does not exist in the same manner as it exists in classical statistics. In particular, the Kolmogorov axioms of statistics are not satisfied. Hence, quantum statistics is a genuine extension of Kolmogorovian statistics.

1.8 The von Neumann Density Operator

The obvious generalisation of the probability distribution of Boltzmann and Gibbs to quantum mechanics is due to *von Neumann*. It assumes a density operator of the form

$$\rho = \frac{1}{Z(\beta)} e^{-\beta H}, \qquad \text{with } Z(\beta) = \operatorname{Tr} e^{-\beta H}. \tag{1.51}$$

Note that $e^{-\beta H}$ is a function of the operator H, not a function of a scalar argument. See the Box 1.4.

(1.51) requires that the operator $e^{-\beta H}$ can be normalised to become a density operator. This is the case if

· the spectrum of the Hamiltonian H is purely discrete

$$H\psi_n = E_n \psi_n, \tag{1.53}$$

with $\psi_n, n = 1, 2, \cdots$ an orthonormal basis of the Hilbert space;
· $\sum_n e^{-\beta E_n} < +\infty$.

This conditions are always fulfilled for quantum systems enclosed in a bounded box. Note however that part of the spectrum of the hydrogen atom is continuous so that the von Neumann density operator does not exist for this system.

Like in the classical case, one can derive an identity by taking the derivative with respect to β of the normalisation condition $\operatorname{Tr} \rho = 1$. One finds

If H is an operator then it is clear what $H^2\psi$ means: one should simply apply the operator H twice on the wave function ψ. In this way one can easily define all powers of H. This gives a first way of defining $\exp(-\beta H)$, namely by the power expansion

$$\exp(-\beta H)\psi = \sum_{n=0}^{\infty} \frac{1}{n!}(-\beta)^n H^n \psi. \tag{1.52}$$

In reality it can be very inconvenient to calculate all powers of H. Therefore, an alternative is needed.

The operators of quantum mechanics are generalisations of what matrices are in the context of finite dimensional spaces. To calculate a function of an Hermitean matrix H, one can first diagonalise the matrix by changing the basis, and then taking the function of the diagonal elements. One can verify easily for powers H^n of H that this gives the same result as applying n times H. Hence, if H is diagonal, with diagonal elements $E_1, E_2, E_3, \cdots$, then $\exp(-\beta H)$ is also diagonal and has diagonal elements $\exp(-\beta E_1), \exp(-\beta E_2), \cdots$.

Box 1.4 Functions of operators

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}\beta}\,\mathrm{Tr}\,\rho &= \frac{\mathrm{d}}{\mathrm{d}\beta}\frac{1}{Z(\beta)}\,\mathrm{Tr}\,e^{-\beta H} \\ &= -\frac{\mathrm{d}}{\mathrm{d}\beta}\ln Z(\beta) - \langle H\rangle. \end{aligned} \tag{1.54}$$

Hence, the same expression for the average energy holds as in the classical case

$$\langle H\rangle = -\frac{\mathrm{d}}{\mathrm{d}\beta}\ln Z(\beta). \tag{1.55}$$

This average energy is often called the *internal energy* and is denoted $U = U(T)$. The *heat capacity,* is the derivative of U

$$C = \frac{\mathrm{d}U}{\mathrm{d}T}. \tag{1.56}$$

As an example, the harmonic oscillator is treated in the Box 1.5.

The Hamiltonian of the quantum mechanical *harmonic oscillator* is

$$H = \frac{1}{2m} P^2 + \frac{1}{2} m\omega_0^2 Q^2, \tag{1.57}$$

with momentum operator P and position operator Q. It has a basis of normalised eigenfunctions $\psi_n(q)$ with corresponding eigenvalues $E_n = (\frac{1}{2}+n)\hbar\omega_0$. In this basis, H is diagonal and satisfies $H\psi_n = E_n\psi_n$. Hence, $\exp(-\beta H)$ is diagonal as well, with eigenvalues

$$e^{-\beta H}\psi_n = e^{-\beta E_n}\psi_n. \tag{1.58}$$

The partition sum can be evaluated explicitly in this basis

$$\begin{aligned} Z(\beta) &= \text{Tr}\,\exp(-\beta H) \\ &= \sum_{n=0}^{\infty} \langle\psi_n|\exp(-\beta H)|\psi_n\rangle \\ &= \sum_{n=0}^{\infty} e^{-\beta E_n} \\ &= e^{-\beta\hbar\omega_0/2} \sum_{n=0}^{\infty} e^{-\beta\hbar\omega_0 n} \\ &= \frac{\sqrt{a}}{1-a} \text{ with } a = e^{-\beta\hbar\omega_0}. \end{aligned} \tag{1.59}$$

The eigenvalues of the density operator ρ are then given by $\rho\psi_n = \lambda_n\psi_n$ with $\lambda_n = (1-a)a^n$. The average energy of the harmonic oscillator equals

$$\begin{aligned} U(T) = \langle H\rangle &= -\frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta) \\ &= -\frac{\mathrm{d}}{\mathrm{d}\beta}\left(-\beta\frac{\hbar\omega_0}{2} - \ln(1-e^{-\beta\hbar\omega_0})\right) \\ &= \hbar\omega_0\left(\frac{1}{2} + \frac{1}{e^{\beta\hbar\omega_0}-1}\right). \end{aligned} \tag{1.60}$$

Box 1.5 The quantum harmonic oscillator

1.9 Fermi-Dirac and Bose-Einstein Distributions

For systems of fermions or bosons the evaluation of the canonical partition sum $Z(\beta)$ is not very easy because the Hilbert space of wave functions is limited to functions which are either anti-symmetric or symmetric under permutation of particles. This restriction makes it difficult to actually perform the summation over a basis of N-particle states. In such cases the grand-canonical ensemble is more convenient.

Important examples of quantum gases are the electrons in a metal or a semiconductor, and the lattice vibrations in a harmonic crystal (called phonons).

The partition function of the grand-canonical ensemble is given by (see (1.34))

$$Z(\beta,\mu) = \sum_{N=0}^{\infty} e^{\beta\mu N} Z_N(\beta), \tag{1.61}$$

where now the N-particle partition sum is given by (1.51). By convention is $Z_0(\beta) = 1$. The parameter μ is the chemical potential. It controls the average number of particles.

In the grand-canonical ensemble the average of an operator A is given by

$$\langle A \rangle = \frac{1}{Z(\beta,\mu)} \sum_{N=0}^{\infty} e^{\beta\mu N} \, \mathrm{Tr}_{\mathcal{H}_N} e^{-\beta H_N} A_N. \tag{1.62}$$

Similar to the classical case the operator A must be defined for each possible number of particles N and is denoted A_N when acting on N-particle wave functions, belonging to $\mathcal{H}_N$.

The grand-canonical partition sum can be calculated explicitly for the *ideal Fermi gas* and for the *ideal Bose gas.* Assume a one-particle Hamiltonian H_1 with a basis of orthonormal eigenfunctions ψ_j and corresponding eigenvalues ϵ_j

$$H_1\psi_j = \epsilon_j\psi_j. \tag{1.63}$$

A basis vector of the Fock space is then determined by specifying numbers $n_1, n_2, \cdots$ where n_j is the number of particles in state ψ_j. In the case of fermions is n_j either 0 or 1, for bosons all non-negative integers are allowed. The total number of particles equals $N = \sum_j n_j$. The energy of the state is $E = \sum_j n_j\epsilon_j$ (it is assumed that the particles do not interact with each other; this is why the model is called *ideal* Fermi/Bose gas). The canonical partition sum is therefore given by

$$Z_N(\beta) = \sum{}' \exp(-\beta \sum_j n_j\epsilon_j) \tag{1.64}$$

where $\sum'$ is the sum over all N-particle states, i.e., over all allowed choices of n_j such that $N = \sum_j n_j$. The evaluation of this sum is usually quite difficult. However, when evaluating the grand-canonical partition sum the constraint $N = \sum_j n_j$ can be removed. Indeed, in the Bose case one has

$$\begin{aligned} Z_{\mathrm{BE}}(\beta,\mu) &= \sum_{N=0}^{\infty} e^{\beta\mu N} \sum{}' e^{-\beta\sum_j n_j\epsilon_j} \\ &= \sum_{N=0}^{\infty} \sum{}' e^{-\beta\sum_j n_j(\epsilon_j-\mu)} \end{aligned}$$

$$\begin{aligned} &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \prod_j e^{-\beta n_j(\epsilon_j-\mu)} \\ &= \prod_j \left(\sum_{n_j=0}^{\infty} e^{-\beta n_j(\epsilon_j-\mu)} \right) \\ &= \prod_j \left(\frac{1}{1-e^{-\beta(\epsilon_j-\mu)}} \right). \end{aligned} \tag{1.65}$$

In the Fermi case the occupation numbers n_j take only the values 0 and 1. Then the partition sum equals

$$\begin{aligned} Z_{\mathrm{FD}}(\beta,\mu) &= \sum_{n_1=0}^{1} \sum_{n_2=0}^{1} \cdots \prod_j e^{-\beta n_j(\epsilon_j-\mu)} \\ &= \prod_j \left(\sum_{n_j=0}^{1} e^{-\beta n_j(\epsilon_j-\mu)} \right) \\ &= \prod_j \left(1+e^{-\beta(\epsilon_j-\mu)} \right). \end{aligned} \tag{1.66}$$

The average number of particles in state j can be obtained from the expression

$$\langle n_j \rangle = -\frac{1}{\beta}\frac{\partial}{\partial \epsilon_j} \ln Z(\beta,\mu). \tag{1.67}$$

One gets

$$\langle n_j \rangle = \mp\frac{1}{\beta}\frac{\partial}{\partial \epsilon_j} \ln \left(1 \pm e^{-\beta(\epsilon_j-\mu)}\right) = \frac{1}{e^{\beta(\epsilon_j-\mu)} \pm 1}. \tag{1.68}$$

The minus sign holds for bosons, and is the famous *Bose-Einstein distribution*. The plus holds for fermions and is known as the *Fermi-Dirac distribution*.

Problems

1.1. Experimental determination of Boltzmann's constant [1]
An *ideal gas*, consisting of N identical non-interacting particles with mass m, is enclosed in an infinitely high half-open cylinder, which is placed in a uniform gravitational field. See the Figure 1.2.

---

[1] I do not know who is the author of this problem. The original experiment was done one hundred years ago by Jean Perrin using a colloidal solution of grains with a diameter of 0.6μm each.

Determine the value of Boltzmann's constant k_B using following experimental data: $g = 10$ m/s^2, $T = 300$ K, $m = 10^{-17}$ kg (the mass of a small polystyrene sphere). The number of particles detected between altitudes h_j and h_{j+1} (measured in micro meters), is denoted n_j, and is given by

h_j	n_j
0	100
25	55
50	31
75	17
100	9

1.2. A quantum spin
The simplest description of the magnetic spin of a particle is by means of the Pauli matrices. These are given by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.69}$$

They satisfy $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbf{I}$ and $\sigma_x \sigma_y = i\sigma_z$ and $\sigma_x\sigma_y + \sigma_y\sigma_x = 0$, and cyclic permutations of these relations.

The Hamiltonian is given by $H = -\mu\sigma_z$. Calculate the average energy of the spin as a function of temperature. Show by explicit calculation that $\langle \sigma_x(t) \rangle = 0$ at all times.

1.3. A quantum particle trapped between two walls
Consider a quantum particle of mass m, freely moving in one dimension between two reflecting walls, say at positions $x = 0$ and $x = L > 0$. The eigenvalues of the Hamiltonian are

$$E_n = \frac{\hbar^2 \pi^2}{2mL^2} n^2, \qquad n = 0, 1, 2, \cdots. \tag{1.70}$$

Calculate the average energy and the heat capacity as a function of temperature in a high temperature approximation.

Hint: Replace the sum over the spectrum by an integration.

Note: This problem treats the quantum ideal gas in dimension 1. Because the problem is too difficult, an approximation is used.

1.4. Two fermions in the canonical ensemble
Calculate the partition sum $Z_2(\beta)$ for two non-interacting identical fermions trapped in a parabolic potential (i.e. $\epsilon_n = \hbar\omega(n + \frac{1}{2})$, $n = 0, 1, \cdots$).

Notes

The present Chapter can be used as an introduction in a second course of statistical physics, or in a first course for more mathematically oriented students.

The contents is fairly traditional, but is usually dispersed over many chapters. By bringing the material together in one chapter the relation between the distinct distributions is clarified.

I learned about the work of the French physicist Jean Perrin (See the Problem 1.1) from Henk Lekkerkerker.

Objectives

1) Classical statistical mechanics

- Have an idea how many molecules are contained in one cubic meter of air.
- Describe a system of classical mechanics in the canonical ensemble.
- Derive the Maxwell distribution from that of Boltzmann-Gibbs.
- Give a practical definition of temperature.
- Describe a classical gas in the canonical ensemble.
- Describe a classical gas in the grand-canonical ensemble.
- Know the definition of the heat capacity in the canonical ensemble.
- Calculate the partition sum of the classical ideal gas in the canonical ensemble.
- Calculate the partition sum of the classical ideal gas in the grand-canonical ensemble.
- Know the ideal gas law.
- Calculate averages by taking derivatives of the logarithm of the partition sum.

2) Quantum statistics

- Know what is a density operator (density matrix).
- Describe a quantum system in the canonical ensemble.
- Calculate the average energy and the heat capacity of a harmonic oscillator as a function of temperature.
- Derive the Bose-Einstein and Fermi-Dirac distributions for an ideal gas.

References

1. Gibbs, J.W.: Elementary principles in statistical mechanics. Reprint. Dover, New York (1960) v, 5, 56, 66

Chapter 2
Statistical Models

2.1 Parameter Estimation in Statistical Physics

Linear regression is the statistical procedure most known by physicists. The Figure 2.1 shows a linear fit on semi-logarithmic paper to the data of the Problem 1.1. Less known is that linear regression is an example of what is called *a model belonging to the exponential family* A definition of this notion follows later on). The parameters of such models can be estimated by measuring quantities, called estimators. In statistical physics these estimators are called *extensive variables* because usually their average values are proportional to the size of the system. The averages of extensive variables are called *extensive parameters*. In contrast, the parameters of the model are called *intensive parameters*.

A well-known model of statistical physics is the *Ising model* on a chain or on a (finite part of a) square lattice. This model also belongs to the exponential family, by construction. It has 2 parameters, the inverse temperature β and the external magnetic field h. Its Hamiltonian, in the case of the chain, is given by

$$H(\sigma) = -J\sum_{n=1}^{N-1}\sigma_n\sigma_{n+1} - h\sum_{n=1}^{N}\sigma_n. \tag{2.1}$$

In this expression the σ_n are stochastic variables that can take on the values ± 1. This Hamiltonian is used to write down a probability distribution

$$p_{\beta,h}(\sigma) = \frac{1}{Z_N(\beta,h)}\exp(-\beta H(\sigma)). \tag{2.2}$$

The normalisation $Z_N(\beta,h)$ is called the *partition sum* and is given by

$$Z_N(\beta,h) = \sum_\sigma \exp(-\beta H(\sigma)). \tag{2.3}$$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_2,

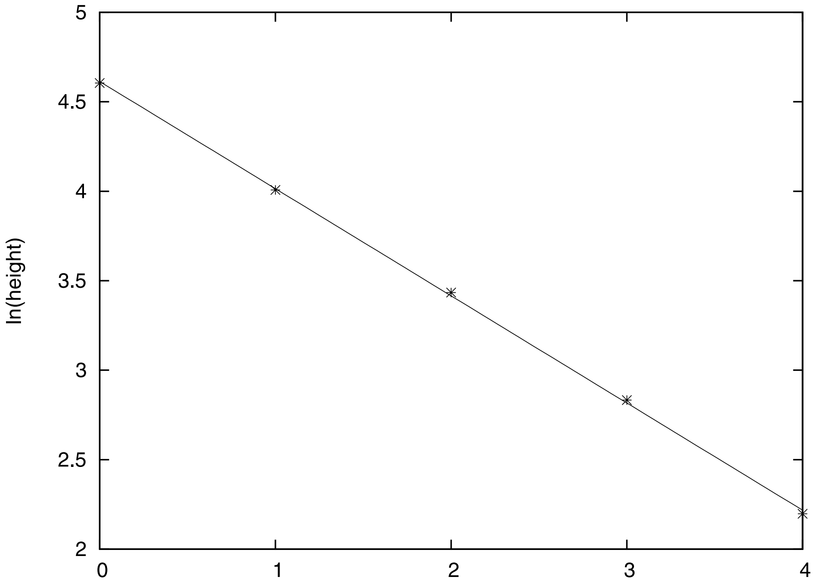

Fig. 2.1 Linear fit to the logarithm of the experimental data of the Problem 1.1

In this expression, the sum $\sum_\sigma$ extends over all possible values of each of the spin variables σ_n.

It is our intention to put the linear regression model and the Ising model on the same footing. To do so requires some work. For the linear regression model we have to introduce a probability distribution $p(y)$, which depends on 3 parameters, the slope a and intercept b of the fitted line, and the root mean square error σ of the fit. We also have to introduce a Hamiltonian H, like that of the Ising model. More precisely, we will introduce 3 extensive variables H_k, one for each of the intensive parameters. Indeed, the Hamiltonian of the Ising model, with its 2 parameters, is the sum of two pieces, one related to the interaction energy, the other due to the external field (here, the size N is not considered as a parameter, although that is a possibility).

Next, the two above mentioned models are considered in the context of statistical parameter estimation. The extensive variables are used as estimators, whose average value can be used to calculate the parameters of the model. In case of linear regression this is clear: the empirical values of the three extensive variables will be used to obtain the fitting parameters a, b, and σ. In case of the Ising model the interaction energy $H_1 = -J\sum_n \sigma_n\sigma_{n+1}$ and the total magnetisation $H_2 = \sum_n \sigma_n$ can be used to estimate the inverse temperature β and the strength of the external field h. The latter looks a little bit strange because experimental measurement of the interaction energy is usually more difficult than measuring temperature. We will come back to this point later on.

2.2 Definition of a Statistical Model

A model, in the present context, is a probability distribution p_θ, depending on a finite number of parameters $\theta^1, \cdots, \theta^m$, together with a set of extensive variables $H_1, \cdots, H_m$, which can be used to estimate the value of the model parameters. The expectation value is denoted $\langle\cdot\rangle_\theta$ and is defined by

$$\langle A\rangle_\theta = \int \mathrm{d}x\, p_\theta(x) A(x). \tag{2.4}$$

In this expression, $A(x)$ is an arbitrary quantity whose value depends on the events x. In the mathematical literature it is called a *stochastic variable*.

We always assume the existence of a function $\Phi(\theta)$ such that

$$\langle H_k\rangle_\theta = -\frac{\partial\Phi}{\partial\theta^k}, \qquad k = 1\cdots m. \tag{2.5}$$

Such a function $\Phi(\theta)$, called a *potential*, exists provided that

$$\frac{\partial}{\partial\theta^l}\langle H_k\rangle_\theta = \frac{\partial}{\partial\theta^k}\langle H_l\rangle_\theta \qquad \text{for all } k, l. \tag{2.6}$$

Its physical meaning is that of a *Massieu function* (this is a kind of *free energy*, well-known in thermodynamics [2]). The precise definition of the Massieu function follows in the next Section.

The estimators H_k are said to be *unbiased* if $\langle H_k\rangle_\theta = \theta^k$. For example, the kinetic energy of a particle in a classical[1] gas, multiplied with an appropriate constant, is an *unbiased estimator* of temperature T. However, most extensive variables of statistical physics are biased estimators.

A quantum model exists of a *density operator*[2] ρ_θ, depending on a finite number of parameters $\theta^1, \cdots, \theta^m$, together with a set of self-adjoint[3] operators $H_1, \cdots, H_m$, which can be used to estimate the value of the model parameters. These operators are the extensive variables of the quantum model. The expectation value of an arbitrary operator A is denoted $\langle\cdot\rangle_\theta$ and is defined by

$$\langle A\rangle_\theta = \operatorname{Tr}\rho_\theta A. \tag{2.7}$$

Characteristic for quantum models is that the operators $H_1, \cdots, H_m$, used to estimate the parameters θ_k, do not necessarily commute between themselves. See the example of Box 2.1.

[1] this means, neglecting quantum effects

[2] see Chapter 1

[3] this is, Hermitean (neglecting some mathematical details)

The simplest quantum-mechanical example concerns a magnetic spin described in terms of *Pauli matrices* $\sigma_k, k = 1, 2, 3$. These are defined by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.8}$$

They satisfy the relations $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \mathbf{I}$, $\sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3$, and cyclic permutations of the latter.

An arbitrary density operator in the Hilbert space $\mathbf{C}^2$ can be written as

$$\rho_{\mathbf{r}} = \frac{1}{2}\left(\mathbf{I} - r^k\sigma_k\right), \tag{2.9}$$

with parameters r_1, r_2, r_3 satisfying $|\mathbf{r}|^2 = \sum_\alpha r_\alpha^2 \leq 1$. See the Figure 2.2. The Pauli matrices are the extensive variables of this model, which is known as the *Bloch representation* of the Pauli spin: $H_k = \sigma_k, k = 1, 2, 3$. A short calculation gives

$$\langle\sigma_k\rangle_{\mathbf{r}} = \frac{1}{2}\,\mathrm{Tr}\,\left(\mathbf{I} - r_l\sigma^l\right)\sigma_k = -r_k. \tag{2.10}$$

The potential $\Phi(\mathbf{r})$ is given by

$$\Phi(\mathbf{r}) = \frac{1}{2}|\mathbf{r}|^2 + \text{ constant.} \tag{2.11}$$

Box 2.1 Quantum spin example

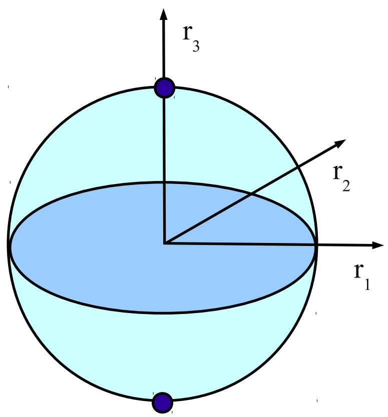

Fig. 2.2 The Bloch sphere. The top point corresponds with the pure state vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, the bottom point with $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

2.3 The Exponential Family

Here we show that the linear regression model belongs to the (curved) exponential family. Let

$$\begin{aligned} \theta^1 &= \tfrac{a}{\sigma^2} & \qquad H_1(y) &= -\textstyle\sum_n x_n y_n \\ \theta^2 &= \tfrac{b}{\sigma^2} & \qquad H_2(y) &= -\textstyle\sum_n y_n \\ \theta^3 &= \tfrac{1}{2\sigma^2} & \qquad H_3(y) &= \textstyle\sum_n y_n^2 \end{aligned}$$

Then one calculates

$$\begin{aligned} \int \mathrm{d}y_1 \cdots \int \mathrm{d}y_N \, e^{-\theta^k H_k(y)} &= \prod_{n=1}^{N} \int \mathrm{d}y_n \, e^{-\frac{1}{2\sigma^2}\left(y_n^2 - 2y_n(ax_n+b)\right)} \\ &= (2\pi\sigma^2)^{N/2} \prod_{n=1}^{N} \exp\left(\frac{1}{2\sigma^2}(ax_n+b)^2\right). \end{aligned}$$

Hence a properly normalised probability distribution $p_\theta(y)$ is obtained when the normalisation is fixed by

$$\begin{aligned} \Phi(\theta) &= \frac{N}{2} \ln 2\pi\sigma^2 + \frac{1}{2\sigma^2} \sum_{n=1}^{N} (ax_n + b)^2 \\ &= -\frac{N}{2} \ln \frac{\theta^3}{\pi} + \frac{1}{4\theta^3} \sum_{n=1}^{N} (\theta^1 x_n + \theta^2)^2. \end{aligned}$$

Three identities are now obtained by taking derivatives

$$\begin{aligned} \langle H_1 \rangle_\theta &= -\frac{\partial \Phi}{\partial \theta^1} = -a \sum_{n=1}^{N} x_n^2 - b \sum_{n=1}^{N} x_n \\ \langle H_2 \rangle_\theta &= -\frac{\partial \Phi}{\partial \theta^2} = -a \sum_{n=1}^{N} x_n - bN \\ \langle H_3 \rangle_\theta &= -\frac{\partial \Phi}{\partial \theta^3} = N\sigma^2 + \sum_{n=1}^{N} (ax_n + b)^2. \end{aligned}$$

These imply the famous fitting formulae

$$\begin{aligned} a &= \frac{1}{N} \frac{\langle x \rangle \langle H_2 \rangle - \langle H_1 \rangle}{\langle x^2 \rangle - \langle x \rangle^2} = \frac{1}{N} \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2} \\ b &= -\frac{1}{N} \langle H_2 \rangle - a\langle x \rangle = \langle y \rangle - a\langle x \rangle. \end{aligned}$$

Box 2.2 Linear regression

The notion of a statistical model has been explained in the previous section. Some statistical models are much easier to analyse than others. This is the case for models belonging to the exponential family. A good understanding of this property, shared by many models, is essential for the present book. It is the corner stone of the first part of the book and is generalised in the second part. The reason why it is so important is that it is the mathematical characterisation of the Boltzmann-Gibbs distribution as it is known in statistical physics – see Chapter 1. The second part of the book deals with generalisations of Boltzmann-Gibbs. These generalised probability distributions are characterised by the property that they belong to a generalised exponential family.

A statistical model is said to belong to the *exponential family* if its probability distribution can be written into the form

$$p_\theta(x) = c(x)\exp(-\Phi(\theta) - \theta^k H_k(x)). \tag{2.12}$$

Note the use of *Einstein's summation convention* (the summation over the index k is implicit). It is essential that the prefactor $c(x)$ and the extensive quantities $H_k(x)$ do not depend on the parameters θ while the normalisation function $\Phi(\theta)$ does not depend on the random variable x. Of course, the prefactor $c(x)$ may not be negative. It plays the role of a *prior probability*, although $\sum_x c(x)$ is not necessarily normalised to one. Therefore, it is a *weight*, rather than a probability distribution. In the physics literature the normalisation $\Phi(\theta)$ is usually written as a prefactor and is then called the *partition sum* $Z(\theta)$. The relation between these functions is $\Phi(\theta) = \ln Z(\theta)$.

It might be necessary to introduce new parameters to bring a statistical model into the canonical form (2.12). Indeed, consider for example the *Poisson distribution*

$$p(n) = \frac{\alpha^n}{n!}e^{-\alpha}, \qquad n = 0, 1, 2, \cdots, \tag{2.13}$$

with parameter $\alpha > 0$. Introduce a new parameter $\theta = -\ln\alpha$. Then the distribution can be written into the form (2.12), with

$$c(n) = \frac{1}{n!} \tag{2.14}$$

$$\Phi(\theta) = \exp(-\theta) \tag{2.15}$$

$$H(n) = n. \tag{2.16}$$

This shows that the Poisson distribution defines a 1-parameter model belonging to the exponential family.

It is clear that the Ising model belongs to the exponential family. In fact, from the definition of the canonical ensemble follows that all its models belong to the exponential family as well. To see that the linear regression model belongs to the exponential family requires some work. See the Box 2.2. An example of a probability distribution not belonging to the exponential family is the *Cauchy distribution*

$$p(x) = \frac{1}{\pi}\frac{a}{x^2 + a^2}, \tag{2.17}$$

where a is a positive parameter. This function is also called a *Lorentzian*.

A nice property of models belonging to the exponential family is that it is easy to calculate the averages $\langle H_k \rangle_\theta$ in terms of the parameters θ. Indeed, from the normalisation condition $1 = \int \mathrm{d}x\, p_\theta(x)$ follows

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta^k} \int \mathrm{d}x\, p_\theta(x) \\ &= \int \mathrm{d}x\, p_\theta(x) \left(-\frac{\partial \Phi}{\partial \theta^k} - H_k(x) \right). \end{aligned} \tag{2.18}$$

This implies

$$\frac{\partial \Phi}{\partial \theta^k} = -\langle H_k \rangle_\theta. \qquad k = 1, \cdots, m. \tag{2.19}$$

This is a well-known formula of statistical physics: extensive parameters are obtained by taking derivatives of the logarithm of the partition sum with respect to control variables. This expression also shows that $\Phi(\theta)$ is the potential function mentioned earlier in (2.6). In the next Chapter on thermodynamics it will be argued that the function $\Phi(\theta)$, as appearing in (2.12), is Massieu's function.

In the linear regression model the empirical values of the estimators H_k are used as a best guess for the average values $\langle H_k \rangle$. Next, (2.19) is used to obtain estimated values of the model parameters. One can wonder whether this is an optimal procedure. This kind of question is addressed in the *maximum likelihood* method. In this approach one poses the question what is the most likely value of the model parameters, given a sample of the total population.

In statistical physics, it is tradition to proceed in a different manner. Models (like the Ising model) can be so complex that most effort goes into the evaluation of $\Phi(\theta)$ as a function of θ. The result is then used to calculate averages $\langle H_k \rangle_\theta$ as a function of the parameters θ_k. This functional dependence is finally compared with experimental results, often in an indirect manner.

2.4 Curved Exponential Families

As said before, a probability distribution of the form

$$p_\zeta(x) = c(x) \exp(-\Phi(\theta) - \theta^k(\zeta) H_k(x)), \tag{2.20}$$

The probability density function of the normal distribution is

$$f_{a,\sigma}(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(u-a)^2\right). \tag{2.21}$$

It can be written as

$$f(u) = \exp\left(-\Phi(\theta) - \theta_1 H_1(u) - \theta_2 H_2(u)\right), \tag{2.22}$$

with

$$\begin{aligned} H_1(u) &= \frac{1}{2}u^2, \quad H_2(u) = u, \quad \theta_1 = \frac{1}{\sigma^2}, \quad \theta_2 = -\frac{a}{\sigma^2}, \\ \Phi(\theta) &= \frac{1}{2}\frac{\theta_2^2}{\theta_1} - \frac{1}{2}\ln(2\pi\theta_1). \end{aligned} \tag{2.23}$$

The curved coordinates are $\sigma = \zeta_1(\theta) = 1/\sqrt{\theta_1}$ and $a = \zeta_2(\theta) = \theta_2/\theta_1$. One verifies that

$$\begin{aligned} \frac{1}{2}\langle u^2\rangle &= \langle H_1\rangle = -\frac{\partial\Phi}{\partial\theta_1} = \frac{1}{2}a^2 + \frac{1}{2}\sigma^2 \\ \langle u\rangle &= \langle H_2\rangle = -\frac{\partial\Phi}{\partial\theta_2} = a. \end{aligned} \tag{2.24}$$

Box 2.3 The normal distribution

involving functions $\theta^k(\zeta)$, is still considered to belong to the exponential family because reparametrisation is allowed. If the transformation $\theta(\zeta)$ is nonlinear then the model with probability distribution p_ζ is said to belong to the *curved exponential family.* A well-known example of the curved exponential family is the *normal distribution,* also called the *Gauss distribution.* See the Box 2.3. Also the linear regression model is curved. See the Box 2.2.

Now, the normalisation condition implies

$$\begin{aligned} 0 &= \frac{\partial}{\partial\zeta^k}\int \mathrm{d}x\, p_\zeta(x) \\ &= \int \mathrm{d}x\, p_\zeta(x)\left(-\frac{\partial\Phi}{\partial\zeta^k} - \frac{\partial\theta^l}{\partial\zeta^k}H_l(x)\right). \end{aligned} \tag{2.25}$$

Hence,

$$\frac{\partial\Phi}{\partial\zeta^k} = -\frac{\partial\theta^l}{\partial\zeta^k}\langle H_l\rangle_\zeta, \tag{2.26}$$

which is not of the form (2.5). If the matrix with components $\dfrac{\partial\theta^l}{\partial\zeta^k}$ is degenerate then this set of equations is underdetermined and it is not possible to obtain the extensive parameters $\langle H_l \rangle_\theta$ as a function of the intensive ζ^k just by solving this set of equations.

2.5 Example: The Ising Model in d=1

It is easy to calculate the partition sum of the Ising chain with $h = 0$. Introduce new stochastic variables $\tau_n = \sigma_n \sigma_{n+1}$. Then the partition sum reads

$$\begin{aligned} Z_N(\beta, h=0) &= 2 \sum_\tau \exp(\beta J \sum_{n=1}^{N-1} \tau_n) \\ &= 2 \prod_{n=1}^{N-1} \sum_{\tau_n = \pm 1} \exp(\beta J \tau_n) \\ &= 2\,(2\cosh(\beta J))^{N-1} . \end{aligned} \tag{2.27}$$

Box 2.4 Partition sum of the Ising chain with $h = 0$

The probability distribution $p_{\beta,h}$ of the one-dimensional *Ising model* is given by (2.2) in terms of the stochastic variables σ_n, $n = 1..N$ (called spin variables). The actual probability space is the set Γ of all configurations. Each configuration assigns the value ± 1 to each of the spin variables. See Figure 2.3 for an example of a configuration with $N = 6$. The normalisation condition is

$$\sum_{x \in \Gamma} p_{\beta,h}(x) = 1. \tag{2.28}$$

The number of configurations is 2^N and increases exponentially with increasing value of N. Hence, one can expect that individual probabilities $p_{\beta,h}(x)$ are very small numbers.

The Ising model, as defined by (2.2), is called the model with *open boundary conditions,* or still, the Ising chain. Its partition sum $Z_N(\beta, h)$ can be

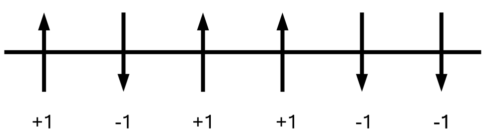

Fig. 2.3 Configuration of spins

Here, we calculate the partition sum of the $d = 1$-Ising model with periodic boundary conditions, using the transfer matrix method.

Write the partition sum as

$$Z_N(\beta, h) = \sum_\sigma \prod_{n=1}^N \exp\left(\beta J \sigma_n \sigma_{n+1} + \frac{1}{2}\beta h(\sigma_n + \sigma_{n+1})\right), \qquad (2.29)$$

where σ_{N+1} is identified with σ_1. Next, notice that this is still identical with

$$Z_N(\beta, h) = \operatorname{Tr} T^N \qquad \text{with} \qquad T = \begin{pmatrix} e^{\beta J + \beta h} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J - \beta h} \end{pmatrix}. \qquad (2.30)$$

The two eigenvalues of this matrix are

$$\lambda_\pm = e^{\beta J} \cosh(\beta h) \pm \sqrt{e^{-2\beta J} + e^{2\beta J} \sinh^2(\beta h)}. \qquad (2.31)$$

The result is

$$Z_N(\beta, h) = \lambda_+^N + \lambda_-^N. \qquad (2.32)$$

In the limit of large N, the so-called *thermodynamic limit,* only the largest eigenvalue is important. The result then simplifies to

$$\ln Z_N(\beta, h) = N \ln \lambda_+ + \cdots. \qquad (2.33)$$

Expansion for small values of h then gives

$$\ln Z_N(\beta, h) = N \ln \cosh(\beta J) + N \ln\left(2 + e^{2\beta J}(\beta h)^2 + \cdots\right). \qquad (2.34)$$

Box 2.5 Partition sum of the $d = 1$-Ising model with periodic boundary conditions

calculated in closed form when $h = 0$. See the Box 2.4. A slight modification of the model allows to calculate $Z_N(\beta, h)$ in closed form for all values of the parameters. Adding one term to the Hamiltonian, (2.1) becomes

$$H(\sigma) = -J \sum_{n=1}^{N-1} \sigma_n \sigma_{n+1} - J\sigma_N \sigma_1 - h \sum_{n=1}^N \sigma_n. \qquad (2.35)$$

The probability distributions, defined by (2.2) with the modified Hamiltonian, are called the Ising model with *periodic boundary conditions,* or, the Ising model on a circle. Its partition sum $Z_N(\beta, h)$ can be calculated in closed form by the *transfer matrix method.* See the Box 2.5.

By taking derivatives of $\ln Z_N(\beta, h)$ with respect to β, respectively h, one obtains the averaged quantities $-\langle H \rangle$ and $\beta\langle M \rangle$, with $M(\sigma) = \sum_{n=1}^N \sigma_n$ the total magnetisation. The resulting expressions are rather complicated. A series expansion for small values of $(\beta h)^2$, together with the approximation that the system size N is large, gives

$$\langle H\rangle = -NJ\tanh\beta J - N(\beta^{-1}+J)e^{2\beta J}(\beta h)^2+\cdots \tag{2.36}$$

$$\langle M\rangle = Ne^{2\beta J}(\beta h)+\cdots. \tag{2.37}$$

Some observations can be made here.

- Both $\langle H\rangle$ and $\langle M\rangle$ are linear in the size of the system N. For this reason, H and M are called extensive variables.
- The average energy $\langle H\rangle$ is a decreasing function of β at constant h. Hence it is an increasing function of the temperature T (the relation between both is $\beta = 1/k_BT$, where k_B is *Boltzmann's constant*; it converts degrees Kelvin into energies, measured in Joule). $\langle H\rangle$ is also called the internal energy. Its derivative with respect to temperature is the heat capacity. A system with negative heat capacity is unstable. Its temperature drops while heating the system. Examples of such behaviour are known in astronomy.
- The average magnetisation $\langle M\rangle$ vanishes when $h = 0$. The derivative of $\langle M\rangle$ with respect to h is called the *static magnetic susceptibility*.
- The Massieu function $\ln Z_N(\beta, h)$ is a real analytic function of $\beta > 0$ and h. The occurrence of a singularity in the function

$$\phi(\beta, h) = \lim_{N\to\infty}\frac{1}{N}\ln Z_N(\beta, h) \tag{2.38}$$

would be associated with a phase transition. These are discussed in a later chapter. The Ising model on a square lattice exhibits such a phase transition.

2.6 The Density of States

Let be given a continuous distribution $p_\theta(x)$ of the form (2.12), belonging to the exponential family. It quite often happens that one is only interested in calculating average values of quantities which depend only on the average values E_k of the Hamiltonians $H_k(x)$. In such a case it is advantageous to introduce the *density of states*

$$\omega(E) = \int \mathrm{d}x\, c(x)\prod_k \delta(H_k(x)-E_k). \tag{2.39}$$

Indeed, the average of a quantity A, which depends only on the $H_k(x)$, is then obtained by

$$\begin{aligned}\langle A\rangle_\theta &= \int \mathrm{d}x\, p_\theta(x)A(H(x))\\ &= \int \mathrm{d}E\,\omega(E)e^{-\Phi(\theta)-\theta^k E_k}A(E).\end{aligned} \tag{2.40}$$

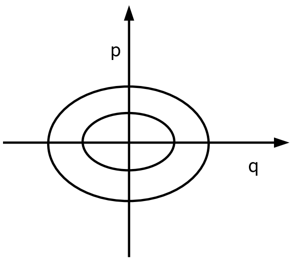

Fig. 2.4 Constant energy lines in the phase space of a harmonic oscillator

Take for example the classical *harmonic oscillator*. Its Hamiltonian is

$$H(x) \equiv H(p,q) = \frac{1}{2m}p^2 + \frac{1}{2}m\omega_0^2 q^2. \tag{2.41}$$

The states of equal energy E in the two-dimensional *phase space* Γ form an ellipse — see the Figure 2.4. Intuitively, one would then guess that the density of states increases with the energy E. However, this is wrong. A careful calculation gives

$$\begin{aligned}
\omega(E) &= \frac{1}{2h}\int \mathrm{d}q \int \mathrm{d}p\, \delta(\frac{1}{2m}p^2 + \frac{1}{2}m\omega_0^2 q^2 - E) \\
&= \frac{1}{h\omega_0}\int \mathrm{d}u \int \mathrm{d}v\, \delta(u^2 + v^2 - E) \\
&= 2\pi \frac{1}{h\omega_0}\int_0^\infty r\mathrm{d}r\, \delta(r^2 - E) \\
&= \pi \frac{1}{h\omega_0}\int_0^\infty \mathrm{d}s\, \delta(s - E) \\
&= \frac{\pi}{h\omega_0}.
\end{aligned} \tag{2.42}$$

This shows that the density of states of the classical harmonic oscillator is a constant, independent of the energy $E \geq 0$.

2.7 The Quantum Exponential Family

A quantum model is said to belong to the *quantum exponential family* if its density operator can be written into the form

$$\rho_\theta = \frac{1}{Z(\theta)}\exp(-\theta^k H_k) = \exp(-\Phi(\theta) - \theta^k H_k), \tag{2.43}$$

with self-adjoint operators H_k and with normalisation

Consider the Bloch representation of the Pauli spin — see the Box 2.1. Let us show that one can write

$$\rho_{\mathbf{r}} = \frac{1}{Z(\mathbf{r})} e^{-\sum_k \theta^k \sigma_k} \tag{2.45}$$

with $\theta^k \equiv \theta^k(\mathbf{r})$ and

$$Z(\mathbf{r}) = \operatorname{Tr} e^{-\theta^k \sigma_k} = 2\cosh|\theta|. \tag{2.46}$$

In particular, this model belongs to the curved quantum exponential family.

In order to prove (2.45, 2.46), chose a basis in which $\rho_{\mathbf{r}} = \frac{1}{2}\left(\mathbf{I} - r^k\sigma_k\right)$ is diagonal. This is equivalent with assuming $r_1 = r_2 = 0$. In that case it is clear that $\rho_{\mathbf{r}} = \frac{1}{Z}\exp(-\theta_3\sigma_3)$ with $\tanh\theta_3 = r_3$ and $Z = 2\cosh\theta_3$. By going back to the original basis $\theta_3\sigma_3$ transforms into a matrix of the form $\theta^k\sigma_k$. The trace of a matrix does not depend on the choice of basis. Hence,

$$Z = 2\cosh\theta_3 = 2/\sqrt{1-\tanh^2\theta_3} = 2/\sqrt{1-r_3^2}. \tag{2.47}$$

Under a change of basis the length of the Bloch vector does not change. Hence, (2.46) follows.

Box 2.6 The Bloch sphere belongs to the quantum exponential family

$$Z(\theta) = \operatorname{Tr}\exp(-\theta^k H_k), \qquad \Phi(\theta) = \ln Z(\theta). \tag{2.44}$$

Note that in the example of the *ideal gas* of bosons or of fermions, discussed in Chapter 1, the relevant observables H_N, N, and n_j, two-by-two commute. However, in general, the operators H_k do not mutually commute, except of course in the one-parameter case. As a consequence, several properties, which hold classically or when the H_k mutually commute, cannot be easily generalised. But thanks to the property called 'cyclic permutation under the trace' the basic relations (2.19), with $\Phi(\theta) = \ln Z(\theta)$, still hold. Indeed, one has

$$\begin{aligned}
\frac{\partial\Phi}{\partial\theta^k} &= \frac{1}{Z(\theta)}\frac{\partial}{\partial\theta^k}\operatorname{Tr} e^{-\theta^l H_l} \\
&= \frac{1}{Z(\theta)}\sum_{n=0}^{\infty}\frac{1}{n!}\frac{\partial}{\partial\theta^k}\operatorname{Tr}(-\theta^l H_l)^n \\
&= \frac{1}{Z(\theta)}\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{j=0}^{n-1}\operatorname{Tr}(-\theta^l H_l)^j(-H_k)(-\theta^l H_l)^{n-1-j} \\
&= \frac{1}{Z(\theta)}\sum_{n=0}^{\infty}\frac{1}{n!}n\operatorname{Tr}(-\theta^l H_l)^{n-1}(-H_k) \\
&= -\frac{1}{Z(\theta)}\operatorname{Tr} e^{-\theta^l H_l} H_k
\end{aligned}$$

$$= -\langle H_k \rangle. \tag{2.48}$$

A reparametrisation of the parameter space is allowed. In this case the expressions become

$$\rho_\zeta = \frac{1}{Z(\theta)} \exp(-\theta^k H_k), \tag{2.49}$$

with self-adjoint operators H_k, with $\theta^k \equiv \theta^k(\zeta)$, and with normalisation

$$Z(\theta) = \operatorname{Tr} \exp(-\theta^k H_k). \tag{2.50}$$

The model is said to be *curved* if the transformation $\theta(\zeta)$ is non-linear. For an example, see the Box 2.1.

Note that one cannot add a prior weight (the $c(x)$ in (2.12) and (2.20)) in the definition of the quantum exponential family because it would spoil the property of cyclic permutation under the trace, essential in the calculation of (2.48).

Problems

2.1. Correlations in the one-dimensional Ising model
Calculate $\langle \sigma_1 \sigma_n \rangle$ for the one-dimensional Ising model with periodic boundary conditions, in absence of an external field (this means $h = 0$). A quantity like $\langle \sigma_1 \sigma_n \rangle$ is an example of a two-point *correlation function* [4]

2.2. The Gamma distribution
The density function of the *Gamma distribution* is given by

$$p(x) = \frac{x^{k-1} e^{-x/b}}{b^k \Gamma(k)}. \tag{2.51}$$

It coincides with the exponential distribution when $k = 1$. Show that the Gamma distribution belongs to the exponential family with two parameters $\theta_1 = 1 - k$ and $\theta_2 = 1/b$.

2.3. Example of the quantum exponential family
Show that the density matrix

$$\rho_\theta = \begin{pmatrix} \theta & 0 \\ 0 & 1-\theta \end{pmatrix}, \qquad 0 < \theta < 1, \tag{2.52}$$

belongs to the curved quantum exponential family.

[4] In fact, $\langle \sigma_1 \sigma_n \rangle - \langle \sigma_1 \rangle \langle \sigma_n \rangle$ is a correlation function. But $\langle \sigma_1 \rangle = \langle \sigma_n \rangle = 0$ holds when $h = 0$.

2.4. Density profile of the earth — See [1].
The mass density of the earth $\rho(r)$ decreases as a function of the distance r to the centre of the earth. Assume a perfect sphere. The radius R, the mass M, and the moment of inertia J are experimentally known. Experimental numbers are $R \simeq 6.36 \times 10^6$ m, $M \simeq 6.0 \times 10^{24}$ kg, and $J \simeq 4.0 \times 10^{37}$ kg m^3. Predict the density at the centre of the earth.

Hints Discretise the density $\rho(r)$ by dividing the sphere into N shells of equal volume $V = 4\pi R^3/3N$. This introduces N variables $\rho_1, \cdots \rho_N$, satisfying

$$H_1(\rho) \equiv \frac{1}{N}\sum_{n=1}^{N} \rho_n = \frac{3M}{4\pi R^3} \tag{2.53}$$

$$H_2(\rho) \equiv \frac{1}{N^{5/3}}\sum_{n=1}^{N} \rho_n (n^{5/3} - (n-1)^{5/3}) = 15J/4\pi R^5. \tag{2.54}$$

Next assume a probability distribution $p_\theta(\rho)$ belonging to the two-parameter exponential family with Hamiltonians $H_1(\rho)$ and $H_2(\rho)$. Fix the parameters θ^1, θ^2 so that $\langle H_1 \rangle = 5568$ kg/m^3 and $\langle H_2 \rangle = 4588$ kg/m^3. Finally, integrate $p_\theta(\rho)$ over all ρ_j but that of the most inner shell to obtain the probability distribution of the latter.

Result The predicted density at the center of the earth is 17140 kg/m^3 (do not take this result too serious!).

2.5. Binomial distribution
Fix an integer $n \geq 2$. The binomial distribution is given by

$$p_a(m) = \binom{n}{m} a^m (1-a)^{n-m}, \quad m = 0, 1, \cdots, n, 0 \leq a \leq 1. \tag{2.55}$$

Show that as a function of the parameter a it belongs to the curved exponential family.

2.6. Weibull distribution
Fix positive parameters k and λ. The *Weibull distribution* is defined on the positive axis by

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}. \tag{2.56}$$

For $k = 1$ this is the *exponential distribution*, for $k = 2$ this is the *Raleigh distribution*. Show that as a function of λ it belongs to the curved exponential family.

Notes

The present chapter is inspired by a paper [3] on parameter estimation in the context of generalised thermostatistics. But most of the contents of this Chapter is fairly standard.

The problem of estimating the state of a quantum system has started only recently — see for instance [4] and the references quoted there.

Objectives

- Explain the notion of an estimator.
- Know how to describe an n-parameter model of statistical physics, both classically and quantum mechanically. Give an example of each of these.
- Give the definition of the exponential family, with and without curvature. Give examples.
- Show that the linear regression model belongs to the exponential family.
- Calculate the average value of an estimator of a model belonging to the exponential family, by taking a derivative of the Massieu function.
- Solve the 1-dimensional Ising model both with open and with periodic boundary conditions.
- Give the definition of the quantum exponential family.

References

1. Bevensee, R.: Maximum entropy solutions to scientific problems. Prentice Hall (1993) 35
2. Callen, H.: Thermodynamics and an Introduction to Thermostatistics, 2nd edn. Wiley, New York (1985) v, 23, 37
3. Naudts, J.: Parameter estimation in nonextensive thermostatistics. Physica A **365**, 42–49 (2006) 36, 162
4. Petz, D., Hangos, K., Magyar, A.: Point estimation of states of finite quantum systems. J. Phys. A **40**, 7955–7970 (2007) 36

Chapter 3
Thermodynamic Equilibrium

3.1 Thermodynamic Configuration Space

Thermodynamics assumes that the experimental measurement of average values is limited to only a few extensive variables. In the example of the Ising model (see the previous Chapter) only two such variables, *total energy* E and *total magnetisation* M are considered. These are the *thermodynamic variables*. We will use the generic notation U_k, $k = 1, \cdots, s$ for them. It is obvious to let coincide these thermodynamic variables with the expectation value $\langle H_k \rangle_\theta$, used before to estimate model parameters θ^k.

The set of values attained by the average values of the thermodynamic variables is the *thermodynamic configuration space*[1]. In the simple setting of two thermodynamic variables, referred to above, the set of values $(\langle E \rangle_\theta, \langle M \rangle_\theta)$ forms a subset of the plane $\mathbf{R}^2$. This subset of the plane is the thermodynamic configuration space.

We assume that the knowledge of the average values U_k suffices to determine the parameters θ^k of the statistical model. Hence, there exists a unique p_θ such that $U_k = \langle H_k \rangle_\theta$. In the present Chapter it is assumed that the probability distributions p_θ of the statistical model coincide with the *thermodynamic equilibrium states*. The notion of thermodynamic equilibrium is discussed further on. It is related to that of *thermodynamic stability*. For further use, let us introduce the space $\mathcal{E}$ of all physically allowed probability distributions. These are the distributions that yield the same expectation of the variables H_k as one of the model distributions p_θ. In formulae this reads

[1] It is common to add entropy S as a coordinate of the thermodynamic configuration space and then to consider in this space a manifold of equilibrium states. See for instance Callen [1], Section 4.2. But then the statement that "each point in the configuration space represents an equilibrium state" is not correct. For that reason we will not consider entropy as a thermodynamic variable but give it a special status as a function of the thermodynamic variables.

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_3,

$$\mathcal{E} = \{p : \text{ there exists } \theta \text{ such that} \\ \sum_{j \in J} p(j) H_k(j) = \langle H_k \rangle_\theta \text{ for all } k\}. \tag{3.1}$$

Of course, the p_θ themselves belong to $\mathcal{E}$, as well as many other probability distributions. Stability concerns the question what happens when the equilibrium distribution p_θ is replaced by some arbitrary probability distribution p in $\mathcal{E}$, reproducing the measured values $U_k = \langle H_k \rangle_\theta$ of the estimators H_k.

3.2 Maximum Entropy Principle

In statistical physics, the *maximum entropy principle,* proposed by Jaynes [4, 5], is a means to select one probability distribution out of the set $\mathcal{E}$ of all physically allowed probability distributions. It relies on the concept of *entropy,* which is a difficult notion, almost 150 years old (Clausius, 1865 — see [6]).

The origin of entropy lies in thermodynamics. It is most known from the second law of thermodynamics, which roughly states that entropy of a closed system can only increase with time. Entropy is now used in many domains of science, not always with the same meaning. Here we will distinguish two definitions, that of *thermodynamic entropy* and that of the entropy functional, which is used in statistical physics, in information theory, as well as in other areas of science.

The *entropy functional* $S(p)$ is a real concave function, defined on the convex set $\mathcal{E}$ of probability distributions p. From these probability distributions p one selects the probability distribution that maximises $S(p)$ under the constraint that the average values $\langle H_k \rangle_p$ have the desired values U_k. This selected p is denoted p^* and is called the *equilibrium state* of the model for the given values U_k of the thermodynamical variables.

Jaynes justified the maximum entropy principle with arguments from information theory. The entropy $S(p)$ can be interpreted as a negative information content of the probability distribution p (Shannon, 1948). See the Box 3.1. Hence the maximum entropy principle proposes to select that probability distribution which has the lowest information content. In this way, no biased information is added to the available information that the $\langle H_k \rangle$ have the desired values U_k.

Shannon's source coding theorem states that in the most efficient way of encoding messages, called entropy encoding, the number of bits used to encode a given message is proportional with the logarithm of $1/p$, where p is the probability with which this message occurs. The less probable a message is, the more information it contains. Hence, $\ln(1/p)$ is a measure for the amount of information contained in the message. If the logarithm with base 2 is used then $\ln_2(1/p)$ is the number of bits that is needed to encode the message. To see this, note that with n bits one can encode 2^n different messages. Giving each of them equal probability $p = 2^{-n}$ the quantity $\ln_2(1/p)$ indeed equals n.

The entropy $S(p) = -\sum_i p_i \ln_2(p_i)$ is then the average amount of information contained in a message i, selected randomly with probability p_i from the set of all possible messages.

Box 3.1 Entropy in information theory

3.3 The Boltzmann-Gibbs-Shannon Entropy Functional

The entropy functional that is most often used is that of Boltzmann and Gibbs. Later on, Shannon used it to establish information theory. The reason for this general use of this entropy functional is that it is intimately related to the exponential family. Indeed, any discrete probability distribution belonging to the exponential family automatically is the solution of a maximum entropy principle involving the *Boltzmann-Gibbs-Shannon (BGS) entropy functional*. See the Box 3.2.

Let us start with a discrete probability distribution. Events $j \in J$ have probabilities $p(j)$. They must be properly normalised

$$\sum_{j\in J} p(j) = 1. \tag{3.5}$$

Then the *entropy functional* $S(p)$ is defined by

$$S(p) = -k_B \sum_{j\in J} p(j) \ln \frac{p(j)}{c(j)}, \tag{3.6}$$

where $c(j)$ is a *weight* satisfying $c(j) > 0$ for all $j \in J$. This weight can be used to encode information which does not depend on the parameters of the model. One often gives the previous definition with $c(j) = 1$. However, in what follows, the more general definition (3.6) is needed.

The function $f(x) = -x \ln x$ is positive in the interval $(0, 1)$. See Figure 3.1. Hence, if $p(j) < c(j)$ for all $j \in J$, then each of the terms in (3.6) is nonnegative. In that case, the sum either converges or diverges to $+\infty$, and the value of the entropy $S(p)$ belongs to the interval $[0, +\infty]$, infinity included.

The fundamental property of any entropy functional is its *concavity*. By definition, this means that for any pair of probability distributions p and q

Theorem 3.1. *Assume that the discrete probability distribution p_θ belongs to a curved exponential family with a weight c and estimators H_k. Then any probability distribution p which satisfies $\sum_{j\in J} p(j)H_k(j) = \langle H_k\rangle_\theta$ for all k, also satisfies $S(p) \leq S(p_\theta)$.*

Proof

Assume that

$$p_\theta(j) = c(j)\exp(-\Phi(\theta) - \zeta^k(\theta)H_k(j)), \qquad j \in J. \tag{3.2}$$

Then one has, denoting $f(x) = -x\ln x$, and using the concavity of $f(x)$,

$$\begin{aligned} S(p) &= \sum_{j\in J} p_\theta(j) f\left(\frac{p(j)}{p_\theta(j)}\right) - \sum_j p(j)\ln\frac{p_\theta(j)}{c(j)} \\ &\leq f\left(\sum_{j\in J} p(j)\right) - \sum_j p(j)\ln\frac{p_\theta(j)}{c(j)}. \end{aligned} \tag{3.3}$$

Note that $f(\sum_{j\in J} p(j)) = f(1) = 0$ and

$$\begin{aligned} -\sum_{j\in J} p(j)\ln\frac{p_\theta(j)}{c(j)} &= \sum_{j\in J} p(j)\left(\Phi(\theta) + \zeta^k(\theta)H_k(j)\right) \\ &= \Phi(\theta) + \zeta^k(\theta)\langle H_k\rangle_\theta \\ &= \sum_{j\in J} p_\theta(j)\left(\Phi(\theta) + \zeta^k(\theta)H_k(j)\right) \\ &= S(p_\theta). \end{aligned} \tag{3.4}$$

Hence, $S(p) \leq S(p_\theta)$ follows.

Box 3.2 Theorem stating that any discrete probability distribution belonging to the (curved) exponential family automatically is the solution of a variational principle

one has

$$S\left(\lambda p + (1-\lambda)q\right) \geq \lambda S(p) + (1-\lambda)S(q), \qquad 0 \leq \lambda \leq 1. \tag{3.7}$$

The Boltzmann-Gibbs-Shannon entropy functional inherits this property from the function $f(x) = -x\ln x$. For a smooth function $f(x)$ to be concave it is enough to check that its second derivative $f''(x) = -1/x$ is negative. The concavity of the entropy functional is fundamental because it implies that by taking averages the entropy cannot decrease. Note that a function is said to be *convex* if minus the function is concave.

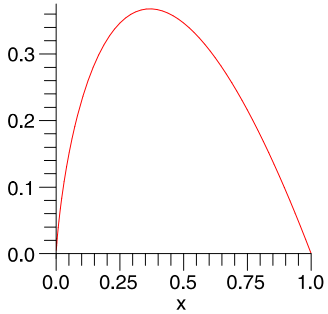

Fig. 3.1 The function $-x \ln x$

3.4 Applying the Method of Lagrange

The direct application of the maximum entropy principle is not very easy. What one needs is a method to find the probability distribution that maximises the entropy functional within a set of distributions all having the right expectations for the thermodynamic variables. Of course, in the case of the BGS-entropy we know (part of) the answer because the probability distributions belonging to the exponential family maximise this entropy. But for other entropy functionals, or, when additional constraints are present, a technique is needed to find the maximising probability distribution.

The *method of Lagrange parameters* is most suited to solve this problem. One introduces Lagrange parameters θ^k, $k = 1, \cdots, s$, and α, and maximises the function

$$\mathcal{L}(p) = S(p) - \theta^k \langle H_k \rangle_p - \alpha \sum_j p(j). \tag{3.8}$$

The parameter α must be chosen in such a way that the normalisation condition $\sum_j p(j) = 1$ is satisfied. The parameters θ^k must be chosen in such a way that the averages $\langle H_k \rangle_p$ have the desired values U_k.

In the case of the BGS-entropy functional one has

$$\mathcal{L}(p) = -\sum_j p(j) \ln \frac{p(j)}{c(j)} - \theta^k \sum_j p(j) H_k(j) - \alpha \sum_j p(j). \tag{3.9}$$

Variation with respect to $p(j)$ yields the condition

$$0 = -\ln \frac{p(j)}{c(j)} - 1 - \alpha - \theta^k H_k(j). \tag{3.10}$$

This can be written into the form

$$p(j) = c(j) \exp\left(-\Phi - \theta^k H_k(j)\right). \tag{3.11}$$

with $\Phi = 1 + \alpha$. Hence, the solution of the optimisation problem found by the method of Lagrange is precisely the exponential family p_θ with parameters θ^k and with estimators H_k.

3.5 Thermodynamic Entropy

The *thermodynamic entropy* is a function $S(U)$, with real values, possibly infinite, defined for points U in the thermodynamic configuration space. It is defined by

$$S(U) = \sup\{S(p) : p \in \mathcal{E} \text{ and } \langle H_k \rangle_p = U_k \text{ for all } k\}. \tag{3.12}$$

The probability distribution p that maximises (3.12), when unique, is the *equilibrium distribution* at the given values of the thermodynamic variables U_k.

In the case that the entropy functional $S(p)$ is that of Boltzmann-Gibbs-Shannon then we know that the distributions p_θ of the exponential family with parameters θ^k and with estimators H_k maximise $S(p)$. Hence, in that case it is true that for each choice of θ^k, if $\langle H_k \rangle_\theta = U_k$, then $S(U) = S(p_\theta)$. In particular, p_θ is the equilibrium distribution in the point U of the thermodynamical configuration space.

It is not very common to express thermodynamic entropy $S(U)$ as a function of energies U_k. Usually, this is done only in the context of what is called the *microcanonical ensemble*, where the conserved quantities H_k have precisely known values. This is not the present context since here we only require in (3.12) that the averages $\langle H_k \rangle_p$ have specific values U_k, while in the microcanonical ensemble one requires that all events j with non-vanishing probability $p(j)$ have values U_k. Instead of working with $S(U)$, one works with its *contact transform,* also called the *Legendre transform*. For doing so, it is important to note that $S(U)$, as defined by (3.12), is automatically a *concave function*. This follows from the concavity of the entropy functional $S(p)$ — see the Box 3.3.

By definition, the Legendre transform of the entropy $S(U)$ is the *Massieu function* $\Phi(\theta)$

$$\Phi(\theta) = \sup_U \{S(U) - \theta^k U_k\}. \tag{3.15}$$

The supremum runs over all points U of the thermodynamic configuration space. The resulting function is automatically convex. Because $S(U)$ is con-

Let U_1 and U_2 be two points in the thermodynamical configuration space, and choose λ in $[0,1]$. For any pair of probability distributions p_1 and p_2, satisfying $\langle H_k\rangle_{p_i} = U_{ik}$, $i = 1, 2$, is

$$\begin{aligned} S(U) &\geq S(\lambda p_1 + (1-\lambda)p_2) \\ &\geq \lambda S(p_1) + (1-\lambda)S(p_2). \end{aligned} \tag{3.13}$$

Because p_1 and p_2 are arbitrary there follows

$$S(U) \geq \lambda S(U_1) + (1-\lambda)S(U_2). \tag{3.14}$$

Box 3.3 Concavity of entropy as a function of energy

cave, the inverse transform

$$S(U) = \inf_\theta \{\Phi(\theta) + \theta^k U_k\} \tag{3.16}$$

yields back the function $S(U)$. For each U in the thermodynamic configuration space for which $S(U)$ is finite there exists usually a unique θ such that the minimum in (3.16) is reached. This correspondence between U and θ defines a function $\theta(U)$. Moreover, it is known from the theory of contact transforms that this function is explicitly given by

$$\theta^k = \frac{\partial S}{\partial U_k}. \tag{3.17}$$

This function $\theta(U)$ solves the estimation problem what parameter values θ^k correspond with the thermodynamic values U_k. Similarly, for each set of parameters θ^k there exists usually a unique point U in the thermodynamic configuration space such that the maximum in (3.15) is reached. Then these functions $U_k(\theta)$ are given by

$$U_k = \langle H_k\rangle_\theta = -\frac{\partial \Phi}{\partial \theta^k}. \tag{3.18}$$

In any case, given a matching pair (U, θ) one has the *thermodynamic relation*

$$S(U) - \Phi(\theta) - \theta^k U_k = 0. \tag{3.19}$$

The relations (3.17, 3.18) together form a pair of *dual identities*.

The Massieu function $\Phi(\theta)$, although of respectable age (1869), is not very well known. The notion used instead is that of *free energy*, which is a contact transform of the energy, not of the entropy. The link between both is usually straightforward. For example, in the *Ising model*, the free energy $F(T, h)$ is given by

$$F(T,h) = U - TS - hM \tag{3.20}$$

while the Massieu function $\Phi(\beta, h)$ is given by

$$\Phi(\beta, h) = S - \beta U + \beta h M. \tag{3.21}$$

In this case, the thermodynamic entropy $S \equiv S(U, M)$ is a function of the energy U and the total magnetisation M. The relations (3.17) become

$$\beta = \frac{1}{T} = \frac{\partial S}{\partial U} \qquad \text{and} \qquad \beta h = -\frac{\partial S}{\partial M}. \tag{3.22}$$

The former equation is often used as definition of the *temperature* T.

3.6 Relative Entropy

Let us from now on assume that the entropy functional $S(p)$ is that of Boltzmann-Gibbs-Shannon. As discussed in the previous Section, this implies that $S(U) = S(p_\theta)$ with parameters θ^k such that $\langle H_k \rangle_\theta = U_k$, and with p_θ belonging to the exponential family. Then one calculates

$$\begin{aligned} S(p_\theta) &= -\sum_j p_\theta(j) \ln \frac{p_\theta(j)}{c(j)} \\ &= -\sum_j p_\theta(j) \left[-\Phi(\theta) - \theta^k H_k(j) \right] \\ &= \Phi(\theta) + \theta^k \langle H_k \rangle_\theta \\ &= \Phi(\theta) + \theta^k U_k. \end{aligned} \tag{3.23}$$

Comparison with the thermodynamic relation (3.19) shows that $\Phi(\theta)$, which enters the definition of the exponential family as the normalisation constant, does indeed coincide with the Massieu function as defined in (3.15).

For any probability distribution p in $\mathcal{E}$ holds that (see the Box 3.2)

$$S(p) \leq S(p_\theta) = S(U) \qquad \text{with } U_k = \langle H_k \rangle_\theta = \langle H_k \rangle_p. \tag{3.24}$$

From the definition (3.15) then follows that

$$\begin{aligned} S(p_\theta) - \theta^k \langle H_k \rangle_\theta &= \Phi(\theta) \\ &\geq S(p) - \theta^k \langle H_k \rangle_p. \end{aligned} \tag{3.25}$$

Hence the quantity $D(p||p_\theta)$, defined by

$$D(p||p_\theta) = (S(p_\theta) - \theta^k \langle H_k \rangle_\theta) - (S(p) - \theta^k \langle H_k \rangle_p) \tag{3.26}$$

cannot be negative. The r.h.s. of (3.25) can be considered to be a non-equilibrium value of the Massieu function, given that the state of the system is described by the probability distribution p instead of the equilibrium distribution p_θ. Then the positivity of (3.26) means that, once the set of parameters θ_k is fixed, the *non-equilibrium Massieu function* is maximal for the equilibrium distribution p_θ.

The expression (3.26) can be written as

$$D(p||p_\theta) = \sum_j p(j) \ln \frac{p(j)}{p_\theta(j)}. \tag{3.27}$$

See the Box 3.4. This is the definition of the *relative entropy* of p with respect to p_θ. In the mathematics literature it is called *divergence*, or also *Kullback-Leibler distance*. However, it is not a distance function in the strict sense, in the first place because in general it is not symmetric under the interchange of its arguments.

Write (3.26) as

$$D(p||p_\theta) = \Phi(\theta) + \sum_j p(j) \ln \frac{p(j)}{c(j)} + \theta^k \sum_j p(j) H_k(j). \tag{3.28}$$

Eliminate $H_k(j)$ from this expression using $\ln p_\theta(j) = -\Phi(\theta) - \theta^k H_k(j)$. This gives the desired result (3.27).

Box 3.4 Derivation of the relative entropy expression

The relative entropy $D(p||p_\theta)$ is a convex function of its first argument p. It vanishes if and only if $p = p_\theta$. These properties will be proved later on in a more general context — see Section 11.2.

3.7 Thermodynamic Stability

The *variational principle* is the statement that the non-equilibrium Massieu function

$$S(p) - \theta^k \langle H_k \rangle_p \tag{3.29}$$

is maximal for the equilibrium distribution p_θ — see the previous Section. More traditionally, as formulated by Gibbs [2, 3], it states that the non-equilibrium free energy is minimal in equilibrium. The physical meaning of the principle is that of *thermodynamic stability* of the equilibrium distribution p_θ. Any perturbation of p_θ leads to a decrease of the Massieu function, or an increase of the free energy. This thermodynamic stability is now rephrased in thermodynamical terms.

Consider a smooth path $t \to \theta(t)$ in the space of model parameters, with initial values $\theta_i^k \equiv \theta^k(t=0)$ and final values $\theta_f^k \equiv \theta^k(t=1)$. A transition of the system from the state described by p_{θ_i} to the state described by p_{θ_f} via the probability distributions $p_{\theta(t)}$ is called a *quasi-stationary process*. The change of entropy between initial and final states can be calculated using the formula

$$S(\theta_f) - S(\theta_i) = \int_{\theta_i}^{\theta_f} \theta^k \mathrm{d}U_k. \tag{3.30}$$

It does not depend on the actual choice of the path $t \to \theta(t)$, but only on its initial and final points. The formal proof of this statement is

$$\mathrm{d}S = \left(\frac{\partial}{\partial\theta^l} S(p_\theta)\right) \mathrm{d}\theta^l = \frac{\mathrm{d}S}{\mathrm{d}U_k}\frac{\partial U_k}{\partial\theta^l}\mathrm{d}\theta^l = \theta^k \mathrm{d}U_k. \tag{3.31}$$

Here, we used (3.17) to evaluate the derivatives of $S(U)$. By integrating this expression, (3.30) follows. Because of (3.31), one says that $\theta^l \mathrm{d}U_l$ is an *exact differential*.

Essential in the above argumentation is the validity of (3.17). More precisely, there should exist a unique tangent plane to the thermodynamic entropy $S(U)$ in each point U of the thermodynamic configuration space. This is automatically the case when p_θ belongs to the exponential family. Indeed, from the properties of the exponential function follows that the functional dependence of $U_k = \langle H_k \rangle_\theta$ can be inverted. The result is the function $\theta^k \equiv \theta^k(U)$. It is then straightforward to see that the derivatives of $S(U) = S(p_\theta)$ with respect to the θ^k, and hence with respect to the U_k exists unambiguously.

Consider now a model belonging to the exponential family and whose single parameter is the inverse temperature β. The *internal energy* U is minus the derivative of the Massieu function $\Phi(\beta)$, and the latter is a convex function. Therefore, the *heat capacity* C satisfies

$$C = \frac{\mathrm{d}U}{\mathrm{d}T} = -\beta^2\frac{\mathrm{d}U}{\mathrm{d}\beta} = \beta^2\frac{\mathrm{d}^2\Phi}{\mathrm{d}\beta^2} \geq 0. \tag{3.32}$$

For this reason, the positivity of the heat capacity is often considered to be a sign of the stability of the model.

3.8 Entropy of Probability Densities

Up to now, the entropy $S(p)$ of discrete probability distributions p has been considered. The obvious generalisation of the BGS-entropy (3.6) to probability densities is

$$S(f) = -\int \mathrm{d}x\, f(x) \ln \frac{f(x)}{c(x)}. \tag{3.33}$$

This expression can be obtained from (3.6) by a limiting procedure. Consider a partition of the configuration space into sets I_j with the property that, up to some epsilon, the function $f(x)/c(x)$ is constant on each of the I_j. Take in each of the I_j an arbitrary point x_j for which $f(x_j) > 0$. Then (3.33) can be approximated as

$$\begin{aligned} S(f) &\simeq -\sum_j \int_{I_j} \mathrm{d}x\, f(x) \ln \frac{f(x_j)}{c(x_j)} \\ &= -\sum_j p(j) \ln \frac{p(j)}{c(j)} \equiv S(p), \end{aligned} \tag{3.34}$$

with

$$p(j) = \int_{I_j} \mathrm{d}x\, f(x) \tag{3.35}$$

$$c(j) = p(j) \frac{c(x_j)}{f(x_j)}. \tag{3.36}$$

These relations are demonstrated in the example of the Box 3.5.

3.9 Quantum Entropies

John von Neumann generalised the Boltzmann-Gibbs entropy functional to the quantum context. The *von Neumann entropy* of a density operator ρ in a Hilbert space $\mathcal{H}$ is defined by

$$S(\rho) = -\operatorname{Tr} \rho \ln \rho, \tag{3.42}$$

and is taken equal to $+\infty$ if $-\rho \ln \rho$ is not *trace class*[2]. Choose an orthonormal basis of eigenfunctions ψ_n of ρ, with corresponding eigenvalues λ_n. Then one has

[2] A positive operator is trace class if its spectrum is purely discrete and the sum of the eigenvalues is finite.

The *exponential distribution* is given by

$$f(x) = ae^{-ax}, \qquad x \geq 0, \tag{3.37}$$

with $c(x) = 1$. Let $x_j = j/n$, $j = 0, 1, \cdots$, for some fixed n. Let $I_j = [x_j, x_{j+1}]$. Then one obtains

$$p(j) = \int_{x_j}^{x_{j+1}} \mathrm{d}x\, f(x) = \left(1 - e^{-a/n}\right) e^{-aj/n} \tag{3.38}$$

$$c(j) = \frac{p(j)}{f(x_j)} = \frac{1}{a}\left(1 - e^{-a/n}\right). \tag{3.39}$$

This discrete distribution p converges to the continuous f as n becomes large. A short calculation gives

$$S(p) = -\ln a + \frac{a}{n}\frac{1}{e^{a/n} - 1}. \tag{3.40}$$

In the limit of large n it converges to

$$S(f) = -\ln a + 1. \tag{3.41}$$

Box 3.5 Entropy of the exponential distribution

$$S(\rho) = -\sum_n \lambda_n \ln \lambda_n \leq +\infty. \tag{3.43}$$

Because all eigenvalues λ_n lie between 0 and 1, and the function $f(x) = -x \ln x$ is positive on the interval $(0, 1)$, the entropy $S(\rho)$ cannot be negative. It shares this property with the discrete probability distributions of the non-quantum case, provided that the weights $c(i)$ all equal 1.

The von Neumann entropy is *concave*. This means that for any pair of density matrices ρ ans σ and for any λ in $[0, 1]$ one has

$$S(\lambda\rho + (1-\lambda)\sigma) \geq \lambda S(\rho) + (1-\lambda)S(\sigma). \tag{3.44}$$

The proof of this statement will be given in a more general context, in Section 11.7.

According to the *maximum entropy principle* the entropy $S(\rho)$ must be maximised given the constraints $\langle H_k\rangle_\rho \equiv \operatorname{Tr} \rho H_k = U_k$. The solution to this optimisation is the quantum exponential family ρ_θ with estimators H_k and parameters $\theta^k = \xi^k$ — See the Box 3.6. The most elegant way to prove this statement involves the *relative entropy*

$$D(\rho||\sigma) = S(\sigma) - S(\rho) - \operatorname{Tr}(\rho - \sigma)\ln\sigma. \tag{3.48}$$

Theorem 3.2. *Let ρ_θ belong to the quantum exponential family with estimators H_k. Assume a density operator ρ satisfies $\mathrm{Tr}\,\rho H_k = \mathrm{Tr}\,\rho_\theta H_k \equiv U_k$ for all k. Then the inequality $S(\rho) \leq S(\rho_\theta)$ holds.*

Proof. One has

$$0 \leq D(\rho||\rho_\theta) = S(\rho_\theta) - S(\rho) - \mathrm{Tr}\,(\rho - \rho_\theta)\ln\rho_\theta. \tag{3.45}$$

Using $\mathrm{Tr}\,\rho H_k = \mathrm{Tr}\,\rho_\theta H_k \equiv U_k$, $\mathrm{Tr}\,\rho = \mathrm{Tr}\,\rho_\theta = 1$, and

$$-\ln\rho_\theta = \Phi(\theta) + \theta^k H_k \tag{3.46}$$

there follows

$$\mathrm{Tr}\,(\rho - \rho_\theta)\ln\rho_\theta = 0. \tag{3.47}$$

Hence, (3.45) implies $S(\rho) \leq S(\rho_\theta)$.

Box 3.6 Quantum variational principle

It satisfies $D(\sigma||\rho) \geq 0$. This inequality is a straightforward consequence of Klein's inequality and is discussed later on in a more general context — see Section 11.2. The difficulty in proving this inequality is that ρ and σ do not necessarily commute. Therefore, there need not exist a basis in which both are simultaneously diagonal.

Problems

3.1. Binomial distribution revisited
The variable x has an integer value n between 0 and N with probability $p(n)$. Known is that $\langle n\rangle = \sum_{n=0}^{N} np(n)$ has the value $\overline{n}$. Calculate the distribution that maximises the BGS-entropy.

Note that if the weights $c(n)$ are given by

$$c(n) = \binom{N}{n} \tag{3.49}$$

then the result is the binomial distribution – see Problem 2.5 of Chapter 2.

3.2. q-deformed distribution
Repeat the previous problem but now with the entropy function

$$S_q^{\rm Tsallis}(p) = \frac{1}{1-q}\left(\sum_n p(n)\left(\frac{p(n)}{c(n)}\right)^{q-1} - 1\right). \tag{3.50}$$

This is Tsallis' entropy – see Chapter 8.

3.3. Entropy in the Bloch representation
The density matrix of a Pauli spin in the Bloch representation is written as (see the Box 2.1)

$$\rho_{\mathbf{r}} = \frac{1}{2}\left(\mathbf{I} - r^k\sigma_k\right). \tag{3.51}$$

Calculate the von Neumann entropy $S(\rho_{\mathbf{r}})$ as a function of $\mathbf{r}$.

3.4. Approximate product measure
Consider two classical spins $\sigma_1 = \pm 1$ and $\sigma_2 = \pm 1$. Try to approximate an arbitrary probability distribution p of the two spin system by a probability distribution q in which the two spins are independent of each other and each have probability a to have the value $+1$ and $1-a$ the value -1. Do this in such a way that the relative entropy $D(p||q)$, which is a kind of distance between p and q, is minimal.

3.5. Maxwell relations
In the constant pressure/constant temperature ensemble one uses the Gibbs potential

$$G(T,p) = \min_{S,V}\{U - TS + pV\}. \tag{3.52}$$

Show that in this ensemble the Maxwell relation $\frac{\partial V}{\partial T} = -\frac{\partial S}{\partial p}$ holds. (Similar relations derived in other ensembles are called Maxwell relations as well).

Notes

The contents of the present Chapter is fairly standard, although the Massieu function is used instead of the free energy because it is the more natural concept. Indeed, in the context of statistical physics and, in particular, of the maximum entropy principle, it is obvious to start from the entropy $S(U)$ as a function of energy U rather than energy $U(S)$ as a function of entropy. The Massieu function $\Phi(\beta)$ is the Legendre transform of $S(U)$ while the free energy $F(T)$ is the Legendre transform of $S(U)$.

The discussion of thermodynamic stability, based on the notion of an exact differential (see (3.31)), goes back to the statement that $\mathrm{d}S = (\mathrm{d}U + p\mathrm{d}V)/T$, with U the internal energy, p the pressure, and V the volume, is an exact differential.

The influence of Edwin Thompson Jaynes (1922-1998) on statistical physics has been tremendous. The maximum entropy principle is the basis for much of the progress both in generalised thermostatistics and in non-equilibrium statistical physics. In the present work it is a theorem rather than an axiom. The notion of relative entropy was introduced in mathematics by Kullback and Leibler who called it the *divergence*. It was used in the 1980's to prove the maximum entropy principle and the variational principle for models of statistical mechanics.

A nice book on the history of thermodynamics has been written by Ingo Müller [6]. A recent discussion of the work of Ludwig Boltzmann has been given by Jos Uffink [7].

Objectives

- Be able to apply the maximum entropy principle to concrete problems.
- Know the BGS-entropy functional and some of its properties.
- Apply the method of Lagrange to optimise the entropy.
- Know the statistical definition of the thermodynamic entropy.
- Explain the use of Legendre transforms.
- Know the relation between the Massieu function and the free energy.
- Know the definition of relative entropy and its relation to the non-equilibrium Massieu function.
- Discuss thermodynamic stability.
- Make the transition between entropy functionals for discrete and for continuous probability distributions.
- Know von Neumann's entropy functional.

References

1. Callen, H.: Thermodynamics and an Introduction to Thermostatistics, 2nd edn. Wiley, New York (1985) v, 23, 37
2. Gibbs, J.W.: On the equilibrium of heterogeneous substances. Transactions of the Connecticut Academy, III. pp. 108–248, Oct., 1875–May, 1876, and 343–524, May, 1877–July, 1878 45
3. Gibbs, J.W.: Scientific papers, Vol. I. Thermodynamics. Longmans, Green, and Co., London, New York and Bombay (1906) 45
4. Jaynes, E.: Information theory and statistical mechanics. Phys. Rev. **106**, 620–630 (1957) 38
5. Jaynes, E.: Papers on probability, statistics and statistical physics, ed. R.D. Rosenkrantz. Kluwer (1989) 38
6. Müller, I.: A History of Thermodynamics. Springer-Verlag, Berlin Heidelberg (2007) 38, 51
7. Uffink, J.: Boltzmann's work in statistical physics. In: Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/statphys-Boltzmann/ (2004) 51

Chapter 4
The Microcanonical Ensemble

4.1 Introduction

The basic concept of this book is the statistical model, which is a parametrised family $p_\theta(x)$ of probability distributions. The present part of the book deals with methods to construct statistical models.

The notion of ensembles was introduced long ago as a tool to derive the fundamentals of statistical physics. This traditional approach is not followed here. But the meaning of the word 'ensemble' is filled in in a way which corresponds with its present day use. In particular, a model treated in the *canonical ensemble* is nothing but a statistical model belonging to the exponential family (see Chapter 2). On the other hand, a model treated in the *microcanonical ensemble* is analysed in the context of mechanics, either classical mechanics or quantum mechanics.

A system in the microcanonical ensemble is considered to be *isolated* from the rest of the world. Of course, such an isolation is an idealisation, which in practice can only be realised on short time scales. In our daily life we encounter mostly systems which are badly isolated. Therefore the results of the microcanonical ensemble may be counter intuitive. But they can be verified experimentally by doing measurements in short enough time so that interaction with the environment through leaks in the isolation can be neglected.

The microcanonical ensemble is important because the validity of a model in the canonical ensemble relies on its mechanical roots and should therefore be derived from the microcanonical ensemble. There are several ways to make such a derivation. This topic is discussed later on. But note that such derivations are easier for classical than for quantum mechanical systems. This is the only point in the present book where the quantum mechanical treatment requires extra care.

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_4,

4.2 The Ergodic Theorem

Consider a system of classical mechanics with N particles. The state of the system is described by coordinates $q_1, \cdots q_N$ and conjugate momenta $p_1, \cdots p_N$. Together they determine a single point q, p in *phase space* Γ — See the Section 1.2. The time evolution $q(t), p(t)$ of the state is determined by the Hamiltonian $H(q,p)$. See the Box 1.1 in Chapter 1.

Of interest are *time averages* of functions $A(q,p)$ that depend on the state q, p of the system. They are defined by

$$\langle A \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{d}t\, A((q(t), p(t)). \tag{4.1}$$

For instance, one would like to know the average kinetic energy of the system. The total energy $H(q,p)$ is a conserved quantity. Hence its time average $\langle H \rangle$ equals $H(q(t), p(t))$ and does not depend on time, but only on the initial conditions. But the kinetic energy fluctuates in time because energy is transferred between the kinetic and the potential energy contributions. Hence, the calculation of the average kinetic energy is a non-trivial problem.

The calculation of time averages is a hard problem because it requires the solution of Hamilton's equations of motion. In addition, the position and momentum as a function of time of each of a large number of particles (typically $N \simeq 10^{22}$ in one litre of air) is such a large amount of information that it is often difficult to handle. Therefore the time average of thermodynamic quantities is seldom calculated. Thanks to the *ergodic theorem* it is possible to replace the time average by an integration over phase space. The latter is exactly what one does in statistical physics — for instance, the Boltzmann-Gibbs distribution (1.12) contains an integration over phase space instead of an integration over time. The ergodic theorem can therefore be seen as the corner stone of statistical mechanics.

The main assumption for the ergodic theorem to hold is the *ergodic hypothesis*. One way of formulating this hypothesis is that one assumes that there exist initial conditions q, p such that the orbit q_t, p_t comes arbitrary close to any point of the phase space Γ. Given this orbit, one then has for any function $A(q,p)$ that

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{d}t\, A((q(t), p(t)) = \int_\Gamma \mathrm{d}q\mathrm{d}p\, A(q,p) \tag{4.2}$$

The proof of this theorem would lead us too far away of our main topic.

The ergodic hypothesis is not often satisfied because of the existence of *conserved quantities*. In particular, the Hamiltonian $H(q,p)$ is a conserved quantity. Hence, an orbit with initial conditions q, p corresponding with a given value of the energy $H(q,p)$ can never come close to a point in phase space with a different value of the energy (assuming that the energy depends continuously on q, p). The solution to this problem is to restrict the phase space Γ to the small part where the relevant conserved quantities are constant and have given values. But, it is not always possible to take all conserved quantities into account. The attitude of statistical physics is then to prefer the average over the phase space above the time average, even if they are not equal to each other. The justification for this attitude is that often the time evolution of a mechanical system is not stable under small perturbations. Adding a little bit of *noise* to the system can have a dramatic effect on time averages while averages over phase space are usually not much affected.

4.3 Example: The Harmonic Oscillator

The classical harmonic oscillator has been considered in Section 2.6. The phase space equals $\Gamma = \mathbf{R}^2$. For any initial condition q, p, different from the ground state $q = p = 0$, the orbit is an ellipse, which is uniquely defined by the value E of the energy $H(q, p)$. Hence, to make the system ergodic, the phase space Γ must be restricted to this ellipse. The average over phase space of a function $A(q, p)$ is then given by

$$\langle A \rangle_E = \frac{1}{2h\omega(E)} \int \mathrm{d}q\mathrm{d}p\, \delta(E - H(q,p))A(q,p), \tag{4.3}$$

with $\omega(E) = \pi/h\omega_0$ the *density of states* — see (2.42). For instance, if A equals the kinetic energy $K = p^2/2m$ this expression can be evaluated

$$\langle K \rangle_E = \frac{1}{2h\omega(E)} \int \mathrm{d}q\mathrm{d}p\, \delta(E - H(q,p))\frac{p^2}{2m} = \frac{1}{2}E. \tag{4.4}$$

Next, let us calculate the time average of the kinetic energy K. The solution of the equations of motion is

$$p(t) = p\cos(\omega_0 t) - m\omega_0 q \sin(\omega_0 t) \tag{4.5}$$

$$q(t) = q\cos(\omega_0 t) + \frac{p}{m\omega_0}\sin(\omega_0 t). \tag{4.6}$$

Hence one obtains

$$K(t) = \frac{1}{2m}\left[p\cos(\omega_0 t) - m\omega_0 q \sin(\omega_0 t)\right]^2. \tag{4.7}$$

The time average of this expression is

$$\langle K \rangle = \frac{1}{2m}\left[\frac{1}{2}p^2 + \frac{1}{2}m^2\omega_0^2\right] = \frac{1}{2}H(p,q) = \frac{E}{2}, \tag{4.8}$$

in agreement with (4.4).

4.4 Definition

The probability distribution of the *microcanonical ensemble*, given values $E_1, \cdots, E_k$ of the conserved quantities $H_1(x), \cdots, H_k(x)$, equals

$$q_E(x) = \frac{c(x)}{\omega(E)} \prod_k \delta(E_k - H_k(x)), \qquad x \text{ in phase space}, \tag{4.9}$$

with normalisation factor

$$\omega(E) = \int \mathrm{d}x\, c(x) \prod_k \delta(E_k - H_k(x)). \tag{4.10}$$

The justification for this probability distribution is that it assigns the same probability to all states that have the right values of the conserved quantities. Correspondingly, the *Boltzmann entropy* of the microcanonical ensemble is defined by

$$\check{\mathcal{S}}(E) = k_B \ln \omega(E). \tag{4.11}$$

Note that $\omega(E)$ is the *density of states*, already introduced in Section 2.6. There it was seen that the density of states of the harmonic oscillator is constant. Hence, also the microcanonical entropy $\check{\mathcal{S}}(E)$ of the harmonic oscillator is constant, this means, does not depend on the energy E.

Note the hacek (or wedge) on top of the symbol $\check{\mathcal{S}}(E)$. Standard textbooks do not put any accent or diacritic on this symbol and identify it with the thermodynamic entropy $S(U)$. However, there is no consensus in the literature that this identification is justified. The problem is related to the question whether one can define the notion of temperature in a microcanonical ensemble. One point of view is that temperature T is defined only in the canonical ensemble where it is the dual parameter of the energy U — see (3.22). On the other hand, it is clear that quite often one can measure experimentally the temperature of an isolated system and that it is desirable to have a definition for the quantity that one measures. And in addition, thermodynamics is valid for isolated systems as well. As mentioned above, the microcanonical entropy $\check{\mathcal{S}}(E)$ of the harmonic oscillator is constant. Identification with $S(U)$, in combination with the thermodynamic definition (3.22) then implies that the temperature of an isolated harmonic oscillator is always infinite instead of being given by the canonical *equipartition* result that the kinetic energy of the harmonic oscillator equals $\frac{1}{2}k_\mathrm{B}T$. Note that the identification of the thermodynamic energy U with the microcanonical E is obvious.

An alternative proposal for the microcanonical entropy, formulated long ago [12, 18, 29, 25], is

$$\mathcal{S}(E) = k_\mathrm{B} \ln \Omega(E) \tag{4.12}$$

where $\Omega(E)$ is the integrated density of states

$$\Omega(E) = \int^{E} \mathrm{d}E\,\omega(E). \tag{4.13}$$

It is immediately clear that $\mathcal{S}(E)$ is an increasing function of the energy E. In addition, the derivative may be interpreted as the inverse temperature β. Indeed, from the definition (4.12) follows

$$\frac{1}{T} \equiv \frac{\mathrm{d}\mathcal{S}}{\mathrm{d}E} = \frac{\omega(E)}{\Omega(E)}. \tag{4.14}$$

For the harmonic oscillator this ratio equals $1/E$, which is the expected result. In fact, it is proved later on that for a gas of N interacting particles the average kinetic energy $\langle K \rangle$ is proportional to the ratio $\Omega(E)/\omega(E)$. More precisely one has always

$$\langle K \rangle = \frac{3N}{2}\frac{\Omega(E)}{\omega(E)} \tag{4.15}$$

so that

$$\left(\frac{\mathrm{d}\mathcal{S}}{\mathrm{d}E}\right)^{-1} = \frac{2}{3N}\langle K \rangle. \tag{4.16}$$

In the classical (i.e. non-quantum) canonical ensemble the equipartition theorem assigns to the kinetic energy a value of $\frac{3}{2}Nk_\mathrm{B}T$. From (4.16) then follows that the alternative definition $\mathcal{S}(E)$ can be identified with the thermodynamic entropy. This seems to settle the problem.

Note that anyhow the difference between the two definitions (4.11) and (4.12 is not extensive (this is, does not increase linearly with the number of particles N) so that the difference between the definitions is not important for large systems.

4.5 Microcanonical Instabilities

A phenomenon known to occur in isolated systems is that of a negative heat capacity. This means that the temperature drops when energy is added to the system, which contradicts our intuition which is based on our extensive experience with badly isolated systems. When the heat capacity is negative then $\beta = 1/k_\mathrm{B}T$ is an increasing function of the energy U. The thermodynamic relation

$$k_\mathrm{B}\beta = \frac{\mathrm{d}S}{\mathrm{d}U} \tag{4.17}$$

then implies that the entropy $S(U)$ is locally convex instead of concave. The thermodynamical requirement of stability is the concavity of the entropy $S(U)$ as a function of energy U — see Chapter 3. More generally, the microcanonical entropy $\mathcal{S}(\eta)$ as a function of microcanonical parameters $\eta_1, \eta_2, \cdots, \eta_n$ must be concave.

In a perfectly isolated system the unstable state of the system may continue to exist. The only way out is to partition the system into two regions, one of low energy and one of high energy. In each of the regions the entropy is concave. This can happen when the energy and entropy of the boundary between the two phases is negligible. Such a separation into two regions with rather different characteristics is called a separation of phases. It is then obvious to characterise a microcanonical *phase transition* by the appearance of a non-concavity of the microcanonical entropy $\mathcal{S}(\eta)$ of the system as a whole.

The non-concavity of the microcanonical entropy $\mathcal{S}(\eta)$ must be related to special properties of the density of states $\omega(\eta)$. Two examples are discussed below. In the example of the Ising model the density of states vanishes in certain regions of the parameter space. When certain states of the system cannot be physically realised then the entropy $\mathcal{S}(\eta)$ cannot be concave in that region. In the example of the pendulum (see below) the density of states diverges at a given value of the energy $E = E_c$. The pendulum clearly has two phases: an oscillatory phase at low energy, and a rotational phase at high energy. The energy value E_c separates these two phases.

The Ising Model

Let us now consider the two-dimensional Ising model in the microcanonical ensemble. We know that this model exhibits a phase transition in the thermodynamic limit. The obvious question is then whether the finite model already shows some indications of this phase transition.

Two configurations of the Ising model have the same energy and the same magnetisation when they have the same number m of upspins and the same number n of nearest neighbour spin pairs with unequal values. The number of such equivalent configurations is denoted $C(m,n)$. For very small systems these counts $C(m,n)$ can be evaluated easily by enumerating all configurations. See the Figure 4.1. In general, these counting numbers have to be estimated by numerical methods. They determine the Boltzmann entropy by

$$\check{\mathcal{S}}(m,n) = k_{\mathrm{B}} \ln C(m,n). \tag{4.18}$$

From Figure 4.1 one notes immediately that the domain where the $C(m,n)$ do not vanish is not convex.

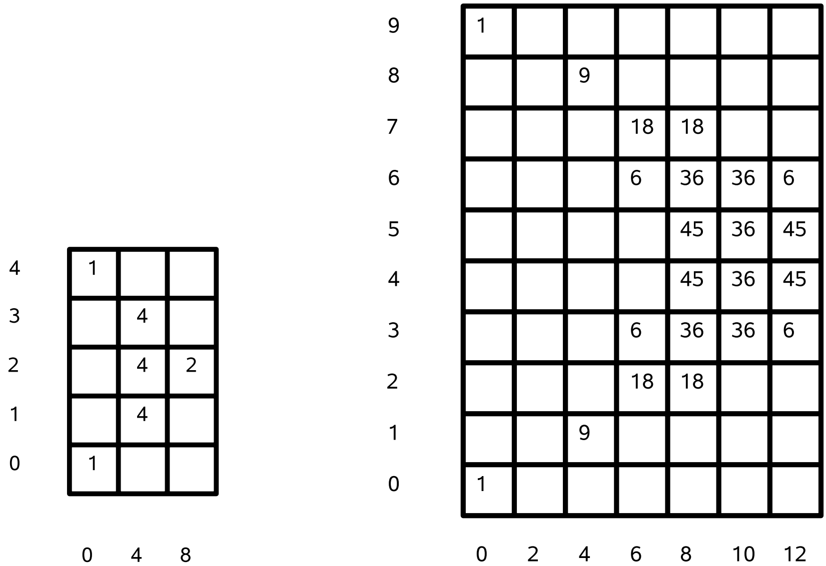

Fig. 4.1 Number of configurations of a 2x2 and a 3x3 Ising lattice as a function of energy and magnetisation. On the vertical axis is the number of upspins m, on the horizontal axis the number of non-matching neighbour pairs n

Example of a Microcanonical Instability

The Hamiltonian of the *pendulum* is

$$H = \frac{1}{2m}p^2 - mk\cos(\phi), \tag{4.19}$$

with $m > 0$ the mass, and with $k > 0$ a constant. As long as the energy E is below mk then the pendulum cannot rotate but librates around $\phi = 0$. At $E = mk$ there is a transition from the librational phase to one of the two rotational phases, rotating either left or right. See the Figure 4.2. Already in the mechanical treatment the system has a bifurcation from one librational to two rotational phases. Hence, one expects a phase transition in the microcanonical ensemble.

The density of states of the pendulum can be calculated analytically. It is given by

$$\omega(E) = \frac{2}{\epsilon}\sqrt{\frac{2}{k}}\,\omega_0(E/km), \tag{4.20}$$

with

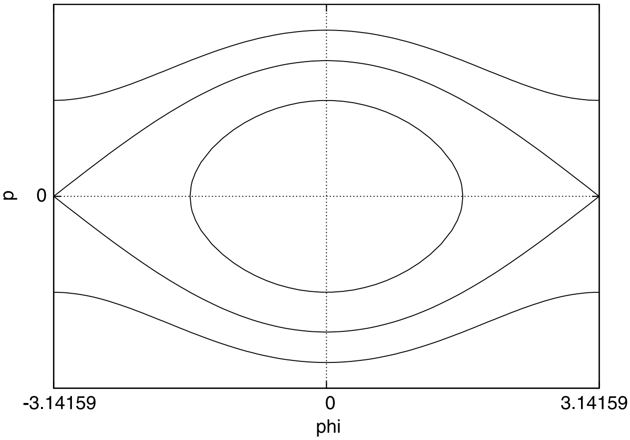

Fig. 4.2 Orbits in the phase space of the pendulum

$$\omega_0(u) = \frac{1}{2\pi} \int_{-\min\{1,u\}}^{1} \mathrm{d}x \, \frac{1}{\sqrt{1-x^2}} \frac{1}{\sqrt{x+u}}. \tag{4.21}$$

See the Figure 4.3. The constant ϵ has been introduced to give $\omega(E)$ the appropriate dimensions. The Boltzmann entropy $\check{\mathcal{S}}(E)$ is piecewise convex instead of concave. The modified entropy $\mathcal{S}(E)$, defined by (4.12), has both concave and convex pieces and is discussed below.

Let us first have a look at the kinetic energy U^{kin} as a function of the total energy E. See the Figure 4.4. Note that it is not a strictly increasing function. When the total energy is slightly below the threshold for rotational motion then the pendulum is very slow on most of its orbit, so that the average kinetic energy is very small. This is a nice example of a negative heat capacity. Slightly increasing the total energy leads to a decrease of the average kinetic energy, which is taken as measure for the temperature of the pendulum.

As a consequence of this unusual behaviour, the free energy $F = U - TS$ is a multi-valued function. See the Figure 4.5. When a fast rotating pendulum slows down due to friction then its energy decreases slowly. The average kinetic energy, which is the temperature $\frac{1}{2}k_{\mathrm{B}}T$, tends to zero when the threshold U_c is approached. In the Figure 4.5, the continuous curve is followed. The pendulum goes from a stable into a metastable rotational state. then it switches to an unstable librating state, characterised by a negative heat capacity. Finally it goes through the metastable and stable librational states. A first order phase transition, directly from the stable rotational phase to the stable librational phase, cannot take place because in a nearly closed system the pendulum cannot get rid of the latent heat. Neither can it stay

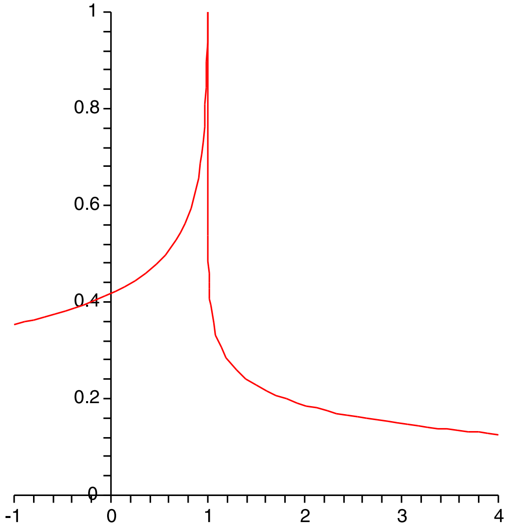

Fig. 4.3 Density of states $\omega_0(u)$ of the pendulum

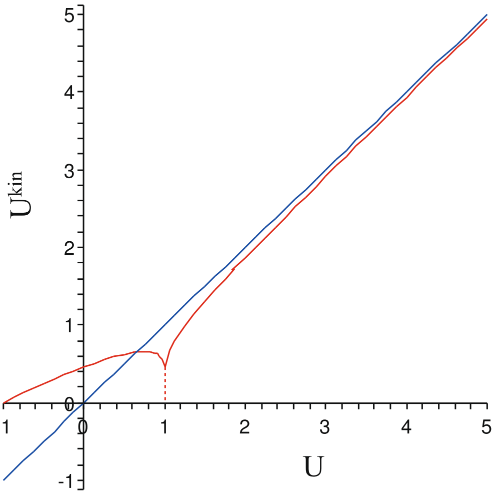

Fig. 4.4 Kinetic energy U^{kin} as a function of the total energy $U = E$

at the phase transition point because a coexistence of the two phases cannot be realised.

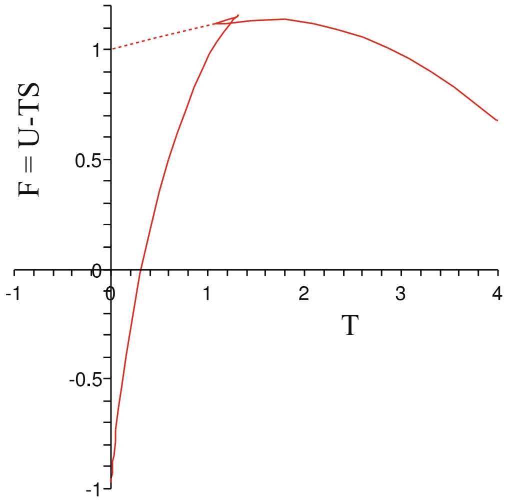

Fig. 4.5 Free energy of the pendulum

4.6 The Quantum Microcanonical Ensemble

In most textbooks the *quantum microcanonical ensemble* is defined as the set of eigenvectors ψ_n of the Hamiltonian H satisfying

$$E \leq E_n < E + \delta, \tag{4.22}$$

where $H\psi_n = E\psi_n$ and $\delta > 0$ is a given small number. Already Erwin Schrödinger in his booklet, dated 1948 [30], remarks that this is not an acceptable definition. There is no reason why the state of an isolated system should be an eigenstate of the energy operator. In fact, it is even not necessarily described by a wavefunction. Indeed, we know that two distinct and non-interacting isolated systems can be described together by a single wavefunction that cannot be written as a product of two wavefunctions, each describing one of the isolated systems. In such a case one says that the two systems are *entangled*. The correct way to describe the state of an isolated system is therefore by means of a density matrix.

An additional difficulty is that quantum mechanics itself gives already a statistical description of the real world. Any density matrix ρ gives automatically rise to a statistical interpretation because its eigenvalues λ_i are positive and normalised $\sum_i \lambda_i = 1$. On top of this quantum probabilities we should add classical probabilities due to our lack of knowledge about the details of the quantum system. In other words, what we need is a distribution function

$f(\rho)$ describing the probability of the density operator ρ. The average of a function $g(\rho)$ is then defined by

$$\langle g \rangle = \int \mathrm{d}\rho\, f(\rho) g(\rho). \tag{4.23}$$

Observables of the quantum system are linear operators. Fix such an operator A. It defines a function $g_A(\rho)$ by the relation $g_A(\rho) = \operatorname{Tr} \rho A$. The average of this function can then be written as

$$\begin{aligned} \langle g_A \rangle &= \int \mathrm{d}\rho\, f(\rho) \operatorname{Tr} \rho A \\ &= \operatorname{Tr} \rho_E A, \qquad \text{with} \quad \rho_E = \int \mathrm{d}\rho\, f(\rho) \rho. \end{aligned} \tag{4.24}$$

Hence, the average over all density operators ρ can be replaced by a calculation involving the average density operator ρ_E.

The Two-level System

The calculation of the density operator ρ_E requires specific information about the ensemble of microcanonical states under consideration. However, in the case of the *two-level atom* there is only one possible candidate for ρ_E. It is uniquely fixed by the requirements that ρ_E is invariant under time evolution and that it predicts the average energy E in a correct manner.

The two-level system is described in terms of *Pauli matrices*. The Hamiltonian is

$$H = -\frac{1}{2} \Delta \sigma_Z. \tag{4.25}$$

The gap between the two energy levels equals Δ. The average density matrix ρ_E should be invariant. Therefore it is of the form

$$\rho_E = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix} = \frac{1}{2} \left(\mathbf{I} + (2p-1)\sigma_z \right). \tag{4.26}$$

From the requirement $\operatorname{Tr} \rho_E H = E$ then follows

$$E = \operatorname{Tr} \rho_E H = -\frac{1}{2}(2p-1)\Delta. \tag{4.27}$$

This fixes the value of p.

The *von Neumann entropy function* (see (3.42)) of ρ_E equals

$$S(E) = -k_{\mathrm{B}} \operatorname{Tr} \rho_E \ln \rho_E$$

$$
\begin{aligned}
&= -k_{\rm B} p \ln p - k_{\rm B}(1-p)\ln(1-p) \\
&= k_{\rm B}\ln 2 - \frac{1}{2}k_{\rm B}(1-\frac{2E}{\Delta})\ln(1-\frac{2E}{\Delta}) \\
&\quad -\frac{1}{2}k_{\rm B}(1+\frac{2E}{\Delta})\ln(1+\frac{2E}{\Delta}).
\end{aligned}
\tag{4.28}
$$

By taking the derivative one obtains an expression for the temperature

$$
\frac{1}{T} = \frac{{\rm d}S}{{\rm d}E} = k_{\rm B}\frac{1}{\Delta}\ln\frac{\Delta - 2E}{\Delta + 2E} \tag{4.29}
$$

This can be written into the familiar form

$$
E = -\frac{\Delta}{2}\tanh\left(\frac{\beta\Delta}{2}\right). \tag{4.30}
$$

Taking the derivative with respect to temperature yields

$$
C = \frac{{\rm d}E}{{\rm d}T} = k_{\rm B}\left(\frac{\beta\Delta}{2}\,{\rm sech}\,\frac{\beta\Delta}{2}\right)^2. \tag{4.31}
$$

At low temperatures, this is, large β, the heat capacity goes exponentially fast to zero. One of the early successes of quantum mechanics was precisely that it explains why the heat capacity of a crystal tends to zero at low temperatures, while the classical *law of Dulong and Petit* predicts that it is constant.

4.7 The Coherent State Ensemble

Another example of a non-conventional microcanonical ensemble is the ensemble of coherent states.

Let H be the Hamiltonian of the quantum harmonic oscillator and let $\psi_n, n = 0, 1, 2, \cdots$ be the eigenstates. They satisfy

$$
H\psi_n = E_n\psi_n \quad \text{with } E_n = \left(\frac{1}{2} + n\right)\hbar\omega_0. \tag{4.32}
$$

The coherent wave functions are of the form

$$
\psi_z(q) = e^{-\frac{1}{2}|z|^2}\sum_{n=0}^{\infty}\frac{1}{\sqrt{n!}}z^n\psi_n(q), \tag{4.33}
$$

where z is an arbitrary complex number. A short calculation gives

$$
\langle H\rangle_z = \langle\psi_z|H|\psi_z\rangle = e^{-|z|^2}\sum_{n=0}^{\infty}\frac{1}{n!}|z|^{2n}E_n = \left(\frac{1}{2} + |z|^2\right)\hbar\omega_0. \tag{4.34}
$$

Hence, the set of wavefunctions ψ_z with $|z|^2 = \frac{E}{\hbar\omega_0} - \frac{1}{2}$ forms a microcanonical ensemble of wave functions with quantum expectation of the energy equal to E.

The time evolution of a coherent wave function can be written down explicitly as

$$e^{-(it/\hbar)H}\psi_z(q) = e^{-\frac{1}{2}|z|^2}\sum_{n=0}^{\infty}\frac{1}{\sqrt{n!}}z^n e^{-itE_n/\hbar}\psi_n(q). \tag{4.35}$$

The expectation of an arbitrary observable A, when averaged over time, becomes

$$\begin{aligned}\langle A\rangle_z &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \mathrm{d}t\,\langle e^{-(it/\hbar)H}\psi_z|A|e^{-(it/\hbar)H}\psi_z\rangle\\ &= e^{-|z|^2}\sum_{n,m=0}^{\infty}\frac{1}{\sqrt{m!n!}}\overline{z}^n z^m\langle\psi_n|A|\psi_m\rangle\lim_{T\to\infty}\frac{1}{T}\int_0^T \mathrm{d}t\,e^{it(E_n-E_m)/\hbar}\\ &= e^{-|z|^2}\sum_{n=0}^{\infty}\frac{1}{n!}|z|^{2n}\langle\psi_n|A|\psi_n\rangle\\ &= \mathrm{Tr}\,\rho_E A\end{aligned} \tag{4.36}$$

with

$$\rho_E = e^{-|z|^2}\sum_{n=0}^{\infty}\frac{1}{n!}|z|^{2n}\,|\psi_n\rangle\,\langle\psi_n|. \tag{4.37}$$

Remember that $|z|^2 = \frac{E}{\hbar\omega_0} - \frac{1}{2}$. Hence, ρ_E depends only on the energy E and not on the choice of initial state ψ_z.

For small E, the result (4.37) can be expanded in series. One obtains

$$\rho_E \simeq (1-|z|^2)|\psi_0\rangle\,\langle\psi_0| + |z|^2|\psi_1\rangle\,\langle\psi_1|. \tag{4.38}$$

This is also the result of the two-level atom discussed before, with $p = 1-|z|^2$. The gap Δ equals $\hbar\omega_0$. Note the shift in the energy scale.

Problems

4.1. Thermodynamic entropy of an ideal gas
a) For an *ideal mono-atomic gas* of N particles of mass m enclosed in a container of volume V the *density of states* equals

$$\omega(E) = \frac{1}{N!h^{3N}\Gamma(3N/2)}V^N(2\pi m)^{3N/2}E^{3N/2-1}. \tag{4.39}$$

Verify this result, using $c(x) = 1/(N!h^{3N})$ in (4.10), and using that the volume of a sphere of radius 1 in dimension n equals $\pi^{n/2}/\Gamma(n/2+1)$. The constant h is arbitrary but has the same units as Planck's constant.
b) Starting from the definition (4.12) of the thermodynamic entropy S, show that it is of the form

$$S \simeq k_{\mathrm{B}} N \left[\frac{3}{2} \ln \frac{2\pi E}{N\epsilon} + \ln \frac{eV}{Na^3} + \text{constant} \right], \tag{4.40}$$

where ϵ and a are constants with units of energy, respectively length. Expression (4.40) is known as the *Sackur-Tetrode equation*. Use Sterling's approximation $\ln \Gamma(z) = (z - 1/2)\ln z - z + \frac{1}{2}\ln 2\pi + \mathrm{o}(z)$.
c) Use this result to derive the equipartition law $E = \frac{3}{2}k_{\mathrm{B}}NT$.

4.2. The quantum harmonic oscillator
Calculate the thermodynamic entropy $S(E)$ of a quantum mechanical harmonic oscillator in the ensemble of the coherent states (see the Section 4.7). Show that it is an increasing and concave function.

Notes

The *ergodic theorem* of Birkhoff dates from 1931, the version of John von Neumann of 1932. However, von Neumann claims that Birkhoff knew about his result when formulating the theorem.

The importance of microcanonical instabilities has been stressed in particular by Dieter Gross [15, 16] and by Alfred Hüller [20].

The use of the integrated density of states $\Omega(E)$ instead of $\omega(E)$, when defining the microcanonical entropy, is usually attributed to Pearson et al [25]. But it was used already in the early works of Gibbs [12] and of Hertz [18], and is treated for instance in the handbook of Becker [3]. It was proposed again [29, 25] in 1948 and in 1985.

The section on the Ising model is partly inspired by [26]. The discussion of the pendulum has been taken from [1]. Other models have been studied as well. In particular, the inequivalence of microcanonical and canonical ensembles has been shown for the infinite-range Blume-Emery-Griffiths model [2]. Hilbert and Dunkel [19] discuss an exactly solvable one-dimensional model. Campa et al have considered a mean field model in the microcanonical ensemble [9, 10, 17]. Further models, treated microcanonically, are the spherical model [4, 23], the Baxter-Wu model and the 4-state Potts model [5].

The definition of quantum statistical averages by integration over wavefunctions has been used recently by Brody and Hughston [8] and by Goldstein et al [14, 13] in the context of the quantum canonical ensemble. Brody et al have considered integration over wavefunctions also for the microcanonical ensemble [8, 7]. Independently, Jona-Lasinio and Prescilla [21, 22] discussed this idea. Other proposals were made in [24, 27], and [11]. However, there is clearly not yet a consensus about the different proposals [24, 6, 28].

An important topic, not discussed here is the *equivalence of ensembles*. A decent treatment of this topic would require a separate Chapter. But the equivalence of ensembles can only hold in the *thermodynamic limit*, which is also not discussed in this book. Hence such a Chapter is out of focus. Note that most textbooks see the canonical ensemble as a Legendre transform of the microcanonical ensemble. This is only correct in the thermodynamic limit because then the canonical probability distribution can be approximated by the microcanonical distribution corresponding to the most likely value of the energy.

Objectives

- Know what the ergodic theorem is about.
- Give the definition of the classical microcanonical ensemble.
- Explain and criticise the definition of Boltzmann's entropy for a model of classical mechanics.
- Discuss the possibility of microcanonical instabilities. Give examples.
- Discuss the two-dimensional Ising model in the microcanonical ensemble.
- Discuss the quantum microcanonical ensemble.

References

1. Baeten, M., Naudts, J.: On the thermodynamics of classical microcanonical systems. arxiv:1009.1787 (2010) 66, 147
2. Barré, J., Mukamel, D., Ruffo, S.: Inequivalence of ensembles in a system with long-range interactions. Phys. Rev. Lett. **87**, 030601 (2001) 66
3. Becker, R.: Theory of Heat. Springer-Verlag, Berlin, Heidelberg, New York (1967) 66
4. Behringer, H.: Critical properties of the spherical model in the microcanonical formalism. J. Stat. Mech. p. P06014 (2005) 66
5. Behringer, H., Pleimling, M.: Continuous phase transitions with a convex dip in the microcanonical entropy. Phys. Rev. E **74**, 011108 (2006) 66
6. Brody, D.: Comment on Typicality for Generalized Microcanonical Ensemble. Phys. Rev. Lett. **100**, 148901 (2008) 67
7. Brody, D., Hook, D., Hughston, L.: Microcanonical distributions for quantum systems. J. Phys. Conf. Ser. **67**, 012025 (2007) 66
8. Brody, D., Hughston, L.: The quantum canonical ensemble. J. Math. Phys. **39**, 6502–6508 (1998) 66
9. Campa, A., Ruffo, S.: Microcanonical solution of the mean field ϕ^4-model: comparison with time averages at finite size. Physica A **369**, 517–528 (2006) 66
10. Campa, A., Ruffo, S., Touchette, H.: Negative magnetic susceptibility and nonequivalent ensembles for the mean field ϕ^4 spin model. Physica A **385**, 233–248 (2007) 66

11. Fine, B.V.: Typical state of an isolated quantum system with fixed energy and unrestricted participation of eigenstates. Phys. Rev. E **80**, 051130 (2009) 67
12. Gibbs, J.W.: Elementary principles in statistical mechanics. Reprint. Dover, New York (1960) v, 5, 56, 66
13. Goldstein, S., Lebowitz, J., Tumulka, R., Zanghi, N.: Canonical typicality. Phys. Rev. Lett. **96**, 050403 (2006) 66
14. Goldstein, S., Lebowitz, J., Tumulka, R., Zanghi, N.: On the distribution of the wave function for systems in thermal equilibrium. J. Stat. Phys. **125**, 1197–1225 (2006) 66
15. Gross, D.: Statistical decay of very hot nuclei, the production of large clusters. Rep. Progr. Phys. **53**, 605–658 (1990) 66
16. Gross, D.: Microcanonical Thermodynamics: Phase transitions in 'small' systems. *Lecture Notes in Physics*, vol. 66. World Scientific (2001) 66
17. Hahn, I., Kastner, M.: Application of large deviation theory to the mean field ϕ^4-model. Eur. Phys. J. B **50**, 311–314 (2006) 66
18. Hertz, P.: Über die mechanischen grundlagen der thermodynamik. Ann. Phys. (Leipzig) **338**, 225–274, 537–552 (1910) 56, 66
19. Hilbert, S., Dunkel, J.: Nonanalytic microscopic phase transitions and temperature oscillations in the microcanonical ensemble: An exactly solvable one-dimensional model for evaporation. Phys. Rev. E **74**, 011120 (2006) 66
20. Hüller, A.: First order phase transitions in the canonical and the microcanonical ensemble. Z. Phys. B **93**, 401–405 (1994) 66
21. Jona-Lasinio, G.: Invariant measures under Schrödinger evolution and quantum statistical mechanics. In: F. Gesztesy, H. Holden, J. Jost, S. Paycha, M. Röckner, S. Scarlatti (eds.) Stochastic Processes, Physics and Geometry: New Interplays. I: A Volume in Honor of Sergio Albeverio. *Canadian Mathematical Society Conference Proceedings*, vol. 28, pp. 239–242 (2000) 66
22. Jona-Lasinio, G., Presilla, C.: On the statistics of quantum expectations for systems in thermal equilibrium. Voluntas International Journal of Voluntary and Nonprofit Organizations **844**, 200 (2006). URL doi:10.1063/1.2219363 66
23. Kastner, M.: Microcanonical entropy of the spherical model with nearest-neighbour interactions. J. Stat. Mech. p. P12007 (2009) 66
24. Naudts, J., der Straeten, E.V.: A generalized quantum microcanonical ensemble. J. Stat. Mech. p. P06015 (2006) 67
25. Pearson, E.M., Halicioglu, T., Tiller, W.A.: Laplace-transform technique for deriving thermodynamic equations from the classical microcanonical ensemble. Phys. Rev. A **32**, 3030–3039 (1985) 56, 66
26. Pleimling, M., Hüller, A.: Crossing the coexistence line at constant magnetization. J. Stat. Phys. **104**, 971–989 (2001) 66
27. Reimann, P.: Typicality for generalized microcanonical ensemble. Phys. Rev. Lett. **99**, 160404 (2007) 67
28. Reimann, P.: Reimann replies. Phys. Rev. Lett. **100**, 148902 (2008) 67
29. Schlüter, A.: Zur Statistik klassischer Gesamtheiten. Z. Naturforschg. **3a**, 350–360 (1948) 56, 66
30. Schrödinger, E.: Statistical thermodynamics: a course of seminar lectures. Cambridge University Press (1948) 62

Chapter 5
Hyperensembles

5.1 Introduction

The primary goal of the present Section is to derive the canonical ensemble from the microcanonical one discussed in the previous Section. The traditional method of constructing ensembles will not be followed. Instead, the recent ideas of superstatistics and of hyperensembles are adapted to suit our purposes. Later on, the approach will be generalised to derive other than canonical ensembles and, in the next Section, to discuss the mean field approximation.

Starting point is a family $q_\eta(x)$ of probability distributions depending on parameters $\eta_1, \eta_2, \cdots, \eta_n$. However, they do not describe the statistical model one is interested in. Instead they are a means to construct the statistical model. How this is done is now explained.

An obvious way to construct a new probability distribution out of the given $q_\eta(x)$ is by assuming that the parameters η are themselves stochastic variables obeying some probability distribution $f(\eta)$

$$p_f(x) = \int \mathrm{d}\eta\, f(\eta) q_\eta(x). \tag{5.1}$$

Hereafter, the distribution $f(\eta)$ is called the *hyperdistribution*. Averages with respect to $p_f(x)$ are denoted

$$\langle A \rangle_f = \int \mathrm{d}x\, p_f(x) A(x). \tag{5.2}$$

Averages with respect to $q_\eta(x)$ are denoted $\langle A \rangle_\eta$ instead of $\langle A \rangle_{q_\eta}$. One has

$$\langle A \rangle_f = \int \mathrm{d}\eta\, f(\eta) \langle A \rangle_\eta. \tag{5.3}$$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_5,

Assume now that the hyperdistribution $f(\eta)$ depends on some parameters $\theta_1, \cdots \theta_k$. Then the distribution $p_f(x)$ is also parametrised by θ and hence can be denoted $p_\theta(x)$. It forms the statistical model that we are interested in. Several choices of hyperdistributions will be discussed. The point is to choose them in such a way that the resulting $p_\theta(x)$ describes a statistical model of interest.

5.2 The Canonical Ensemble

The micro-canonical ensemble, discussed in the previous Chapter, is the corner stone of statistical mechanics because it makes the link with the underlying mechanical theory, either classical mechanics of quantum mechanics. However, it requires the exact knowledge of a number of conserved quantities such as the total energy U. In reality, the total energy is usually not known. One rather measures the temperature, which is only an indirect monitor for the total energy. For this reason, one often works in the canonical or the grand-canonical ensemble. In addition, micro-canonical calculations are known to be more difficult than canonical or grand-canonical ones.

The probability distribution $q_E(x)$ of the micro-canonical ensemble is given by (4.9). Assume now a hyperdistribution $f(E)$. Then the statistical model reads

$$\begin{aligned} p_f(x) &= \int \mathrm{d}E\, f(E) q_E(x) \\ &= c(x) \int \mathrm{d}E\, \frac{f(E)}{\omega(E)} \delta(E - H(x)) \\ &= c(x) \frac{f(H(x))}{\omega(H(x))}. \end{aligned} \tag{5.4}$$

The distribution $f(E)$ of the energy E is unknown. But it is obvious to use the *maximum entropy principle* to fix $f(E)$ under the constraint that the average energy has a given value U. The entropy to be considered here contains two contributions: the microcanonical entropy $\mathcal{S}(E)$ and a contribution due to the lack of information about the actual value of the energy. For the latter, it is obvious to use the *Boltzmann-Gibbs-Shannon entropy functional*

$$S^{\mathrm{BG}}(f) = -k_{\mathrm{B}} \int \mathrm{d}E\, f(E) \ln \frac{f(E)}{m(E)}. \tag{5.5}$$

Note the prior weight $m(E)$ which has been inserted.

The optimisation involves a *Lagrange parameter* β controlling the average energy, and a parameter α needed to ensure the normalisation of the distribution function $f(E)$. The quantity to be optimised is

$$\mathcal{L} = S^{\mathrm{BG}}(f) + \int \mathrm{d}E\, f(E)\mathcal{S}(E) - k_{\mathrm{B}}\alpha \int \mathrm{d}E\, f(E) - k_{\mathrm{B}}\beta\langle H\rangle_f. \quad (5.6)$$

Variation with respect to $f(E)$ gives

$$0 = -\ln\frac{f(E)}{m(E)} - k_{\mathrm{B}} - \alpha + \mathcal{S}(E) - \beta E. \quad (5.7)$$

Hence, one obtains the optimal choice

$$f(E) = \frac{m(E)}{Z(\beta)} e^{\mathcal{S}(E)/k_{\mathrm{B}} - \beta E}, \quad (5.8)$$

with

$$Z(\beta) = \int \mathrm{d}E\, m(E) e^{\mathcal{S}(E)/k_{\mathrm{B}} - \beta E}. \quad (5.9)$$

The resulting family of probability distributions $p_\beta(x)$ is then

$$p_\beta(x) = \frac{c(x)}{Z(\beta)}\frac{m(E)}{\omega(E)} e^{\mathcal{S}(E)/k_{\mathrm{B}}}\, e^{-\beta E}\bigg|_{E=H(x)}. \quad (5.10)$$

This reduces to the Boltzmann-Gibbs distribution (1.12) provided that we make the choice

$$m(E) = \omega(E) e^{-\mathcal{S}(E)/k_{\mathrm{B}}}. \quad (5.11)$$

If now $\mathcal{S}(E)$ equals the Boltzmann entropy $\check{\mathcal{S}}(E) = k_B \ln\omega(E)$ (see (4.11)) then the prior weight $m(E)$ equals 1, which is what one expects. In the case of the modified entropy $\mathcal{S}(E)$, given by (4.12), the weight $m(E)$ is not constant and depends on the system at hand, which is unnatural. Of course it is also possible to defend the point of view that $m(E) \equiv 1$, and that $\mathcal{S}(E)$ differs from the Boltzmann entropy $\check{\mathcal{S}}(E) = k_B \ln\omega(E)$, but that the Boltzmann-Gibbs distribution (1.12) is only exact in the limit of a large system. Some evidence for the latter point of view will be given later on.

5.3 Superstatistics

Before generalising the construction made in the previous Sections, let us digress for a moment into superstatistics.

Given a Hamiltonian $H(x)$, the obvious choice for the probability distributions $q_\eta(x)$ is the *Gibbs distribution* — see (1.12).

$$q_\beta(x) = \frac{c(x)}{Z(\beta)} e^{-\beta H(x)}. \quad (5.12)$$

Up to now, the inverse temperature β was always considered as one of the parameters θ^k of a statistical model. In superstatistics, one considers the possibility that the model parameter β is not constant but is itself a random variable which has some probability distribution. To justify this point of view one refers to inhomogeneous systems out of equilibrium, where the temperature is not a global constant but varies spatially over macroscopic distances. One can then establish experimentally a distribution function $f(\beta)$ on $[0,+\infty)$, which indicates the probability to find a given inverse temperature β at a given spot in the system. Such macroscopic fluctuations have been studied for instance in the context of turbulence — see [4].

Introduce the Laplace transform of the ratio $f(\beta)/Z(\beta)$

$$F(z) = \int_0^{+\infty} \mathrm{d}\beta\, \frac{f(\beta)}{Z(\beta)} e^{-\beta z}. \tag{5.13}$$

Then the probability distribution $p_f(x)$ of the hyperensemble becomes

$$\begin{aligned} p_f(x) &= \int_0^{\infty} \mathrm{d}\beta\, f(\beta) q_\beta(x) \\ &= c(x) F(H(x)). \end{aligned} \tag{5.14}$$

Consider for instance a variable $A(E)$ which depends only on energy E and let $\rho(E)$ be the density of states introduced by (2.39). Then the superstatistical average of A equals

$$\begin{aligned} \langle A \rangle_f &= \int \mathrm{d}x\, p_f(x) A(H(x)) \\ &= \int \mathrm{d}E\, \rho(E) F(E) A(E). \end{aligned} \tag{5.15}$$

See the Box 5.1 for an example.

Because $f(\beta)$ and $Z(\beta)$ are both positive, all derivatives $F^{(k)}(z)$ of $F(z)$ exist and have alternating signs. A positive function $F(z)$ satisfying

$$(-1)^k F^{(k)}(z) \geq 0 \qquad \text{for } z > z_0, k = 0, 1, 2, \cdots \tag{5.22}$$

is said to be *completely monotonic* on the interval $(z_0, +\infty)$. By the theorem of Bernstein, such a function is the Laplace transform of a positive function. Hence, if $p(x)$ is any probability distribution which can be written into the form

$$p(x) = c(x) F(H(x)) \tag{5.23}$$

for some completely monotonic function $F(z)$, then there exist a hyperdistribution f such that $p(x) = p_f(x)$. One concludes that all probability distributions of superstatistics are characterised in this way. Hereafter, these probability distributions will be called *superstatistical distributions*.

Let the hyperdistribution $f(\beta)$ on $[0, +\infty)$ be given by $f = f_a$ with $a > 0$ and

$$f_a(\beta) = \frac{\beta}{a^2} e^{-\beta/a}. \tag{5.16}$$

With this choice of function f the probabilities $p_f(x)$ depend again on a single parameter $a > 0$.

The partition sum of a classical *harmonic oscillator* with frequency ω equals

$$f(q,p) = \frac{1}{Z(\beta)} e^{-\beta H(q,p)}, \tag{5.17}$$

with

$$H(q,p) = \frac{1}{2m} p^2 + \frac{1}{2} \omega^2 q^2 \quad \text{and} \quad Z(\beta) = 2\pi / h\beta\omega. \tag{5.18}$$

Then the function $F_a(z)$, as given by (5.13), is found to be

$$\begin{aligned} F_a(z) &= \int_0^{+\infty} \mathrm{d}\beta \, \frac{h\beta\omega}{2\pi} \, \frac{\beta}{a^2} e^{-\beta/a} e^{-\beta z} \\ &= \frac{ha\omega}{\pi} \frac{1}{(1+az)^3}. \end{aligned} \tag{5.19}$$

The resulting family of probability distributions is

$$p_a(q,p) = \frac{ha\omega}{\pi} \frac{1}{(1 + aH(q,p))^3}. \tag{5.20}$$

Using (5.15), the expression for the superstatistical average energy becomes

$$\begin{aligned} \langle H \rangle_a &= \int_0^{+\infty} \mathrm{d}E \, \frac{\pi}{h\omega} \, \frac{ha\omega}{\pi} \, \frac{1}{(1+aE)^3} E \\ &= \frac{1}{2a}. \end{aligned} \tag{5.21}$$

Box 5.1 Simple example of a superstatistical distribution

Of course, the Boltzmann-Gibbs distribution (5.12) is of the form (5.23), with $F(z) = e^{-\beta z}/Z(\beta)$. Hence, the notion of superstatistical distribution is a generalisation of the notion of Boltzmann-Gibbs distribution.

Not all probability density functions have a completely monotonic Laplace transform. Examples of such probability distributions will follow in later chapters.

5.4 The Hyperensemble

In superstatistics, the hyperdistribution $f(\eta)$ is determined by the physical problem that one wants to model. In a more general context the distribution of the parameters η may be unknown. It is then obvious to use the *maximum entropy principle* to determine the hyperdistribution $f(\eta)$. The resulting family of probability distributions $p_\theta(x)$ will then be called the *hyperensemble* of the family $q_\eta(x)$. With this terminology, the canonical ensemble is the hyperensemble of the microcanonical ensemble. In the quantum case, the hyperensemble of the ensemble of coherent states coincides with the quantum canonical ensemble, but only for linear functions of the density operator, this is, for functions of the form $g_A(\rho) = \operatorname{Tr}\rho A$.

Because the lack of knowledge about the hyperdistribution $f(\eta)$ is a problem of information theory it is indicated to use the Boltzmann-Gibbs-Shannon entropy

$$S(f) = -\int \mathrm{d}\eta\, f(\eta) \ln \frac{f(\eta)}{m(\eta)} \tag{5.24}$$

for the entropy contribution due to the uncertainty about the parameters η. Note the prior weight $m(\eta)$ has been inserted.

The choice of the constraints is crucial because it determines the resulting statistical model. We want to allow a finite number of parameters $\theta_1, \cdots, \theta_k$. Correspondingly, we need Hamiltonians $H_1, \cdots, H_k$ which can be used to estimate these parameters θ_j. In addition, one can take into account that the distributions $q_\eta(x)$ are weighed by some entropy functional $\check{\mathcal{S}}(\eta)$, which for the moment is not further specified. Then, the quantity to be optimised is

$$\mathcal{L} = S(f) - \alpha \int \mathrm{d}\eta\, f(\eta) + \int \mathrm{d}\eta\, f(\eta)\check{\mathcal{S}}(\eta) - \theta^k \langle H_k \rangle_f. \tag{5.25}$$

Variation with respect to $f(\eta)$ gives

$$0 = -\ln \frac{f(\eta)}{m(\eta)} - 1 - \alpha + \check{\mathcal{S}}(\eta) - \theta^k \langle H_k \rangle_\eta. \tag{5.26}$$

This can be written as $f(\eta) = f_\theta(\eta)$ with

$$f_\theta(\eta) = \frac{m(\eta)}{Z(\theta)} \exp\left(\check{\mathcal{S}}(\eta) - \theta^k \langle H_k \rangle_\eta\right), \tag{5.27}$$

with $Z(\theta)$ the appropriate normalisation constant. This $f_\theta(\eta)$ is the *hyperdistribution* of the *hyperensemble*, given the Lagrange parameters $\theta_1, \cdots, \theta_k$.

The hyperdistribution $f_\theta(\eta)$ belongs to the exponential family. Hence, identities are obtained by taking derivatives of the logarithm of the partition sum $Z(\theta)$. One obtains

$$\frac{\partial}{\partial\theta^k}\ln Z(\theta) = -\langle H_k\rangle_\theta. \tag{5.28}$$

This shows that

$$\Phi(\theta) = \ln Z(\theta) \tag{5.29}$$

is the *Massieu function* of the hyperensemble. One then has

$$\frac{\partial\Phi}{\partial\theta^k} = -\langle H_k\rangle_\theta. \tag{5.30}$$

Note that

$$\frac{\partial^2\Phi}{\partial\theta^k\partial\theta^l} = \langle (H_k - \langle H_k\rangle_\theta)(H_l - \langle H_l\rangle_\theta)\rangle_\theta. \tag{5.31}$$

This implies that the matrix of second derivatives has positive eigenvalues. Hence $\Phi(\theta)$ is a convex function, as it should be.

5.5 Properties of the Hyperensemble

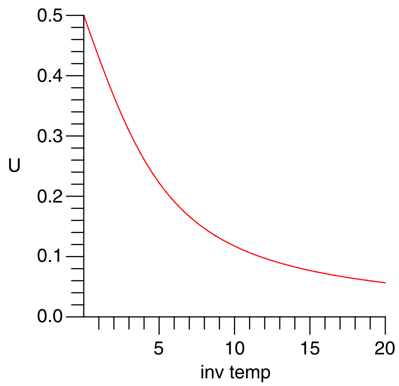

Fig. 5.1 Average energy as a function of inverse temperature for the two-state model

Let us assume that for each vector U there exists a unique set of parameters θ such that

$$U_k = \langle H_k\rangle_\theta. \tag{5.36}$$

The coin tossing experiment has two outcomes, tail and head. These are numbered 0 and 1 and have probabilities $1-\eta$ and η, with $0 \le \eta \le 1$. Let $q_\eta(0) = 1-\eta$ and $q_\eta(1) = \eta$. The Hamiltonian corresponding with this unique parameter η is chosen to satisfy $H(0) = 0$ and $H(1) = 1$. One has

$$\check{\mathcal{S}}(\eta) \equiv S(q_\eta) = -\eta \ln \eta - (1-\eta)\ln(1-\eta) \tag{5.32}$$

$$\langle H \rangle_{q_\eta} = \eta. \tag{5.33}$$

Hence, the hyperdistribution equals (using β as parameter instead of θ)

$$f_\beta(\eta) = \frac{1}{Z(\beta)} \exp\left(-\eta \ln \eta - (1-\eta)\ln(1-\eta) - \beta\eta\right), \tag{5.34}$$

with

$$Z(\beta) = \int_0^1 \mathrm{d}\eta \, \exp\left(-\eta \ln \eta - (1-\eta)\ln(1-\eta) - \beta\eta\right). \tag{5.35}$$

The evaluation of $U = \langle H \rangle_\beta$ has to be done numerically. See the Figure 5.1. It is a monotonically decreasing function, as expected. There is a one-to-one mapping between U and β for $0 < U < 0.5$. The model is thermodynamically stable because the entropy increases with increasing energy and the energy U increases with increasing temperature.

Box 5.2 The two-state example

Then one can define the *thermodynamic entropy* $S(U)$ by

$$S(U) = S(f_\theta) + \int \mathrm{d}\eta \, f_\theta(\eta)\check{\mathcal{S}}(\eta). \tag{5.37}$$

It has two contributions. The average of the entropy $\check{\mathcal{S}}(\eta)$ is augmented with the entropy $S(f_\theta)$ due to the uncertainty about the value of the parameters η.

From (5.27) follows

$$S(f_\theta) = \Phi(\theta) - \int \mathrm{d}\eta \, f_\theta(\eta)\check{\mathcal{S}}(\eta) + \theta^k \langle H_k \rangle_\theta. \tag{5.38}$$

Hence one has the *thermodynamic relation*

$$S(U) = \Phi(\theta) + \theta^k U_k. \tag{5.39}$$

Using (5.30) one obtains

$$\frac{\partial S}{\partial U_l} = \theta^l + \frac{\partial \Phi}{\partial U_l} + \frac{\partial \theta^k}{\partial U_l} U_k$$

$$= \theta^l + \frac{\partial \theta^m}{\partial U_l}\left[\frac{\partial \Phi}{\partial \theta^m} + \frac{\partial \theta^k}{\partial \theta^m} U_k\right]$$
$$= \theta^l. \tag{5.40}$$

This is the dual relation of (5.30). It generalises (3.17). In the case of one parameter this implies that θ equals the *inverse temperature* β, as defined by (3.22).

In the previous Chapter it was shown that $S(U)$ and $\Phi(\theta)$ are each others *Legendre transforms*. The same result holds here for hyperensembles. Indeed, because $\Phi(\theta)$ is convex one has for all η that

$$S(U) = \Phi(\theta) + \theta^k U_k \leq \Phi(\eta) + \eta^k U_k. \tag{5.41}$$

To see this, note that

$$g(\eta) = \Phi(\theta) + U_k(\eta^k - \theta^k) \tag{5.42}$$

defines a plane tangent to the convex $\Phi(\eta)$ surface. One has therefore

$$S(U) = \inf_\theta \{\Phi(\theta) + \theta^k U_k\}. \tag{5.43}$$

This shows that $S(U)$ is the Legendre transform of $\Phi(\theta)$. A well-known consequence is that $S(U)$ is a concave function. The inverse relation

$$\Phi(\theta) = \sup_U \{S(U) - \theta^k U_k\} \tag{5.44}$$

is automatically satisfied because $\Phi(\theta)$ is convex.

See the Box 5.2 for an example.

Notes

The notion of *superstatistics* is due to Christian Beck and Eddie Cohen [2, 3]. Further references are [1, 7]. The notion of *hyperensembles* has been introduced by Gavin Crooks [5] as a tool for describing systems out of equilibrium. Part of the Chapter is an adaptation of results taken from the latter paper.

Up to now, superstatistics has been discussed almost exclusively in a non-quantum context. A quantum generalisation was proposed by A.K. Rajagopal [6]. In his approach the hyperdensity distribution $f_\theta(\eta)$ is replaced by a *positive operator-valued measure*.

Objectives

- Describe the probability distribution of superstatistics using Laplace transforms.
- Characterise the probability distributions of superstatistics by means of the theorem of Bernstein.
- Derive the hyperdistribution of a hyperensemble using the maximum entropy principle.
- Show that the thermodynamic entropy $S(U)$ and the Massieu function $\Phi(\theta)$ of a hyperensemble are related by a Legendre transform.

References

1. Abe, S., Beck, C., Cohen, E.G.D.: Superstatistics, thermodynamics, and fluctuations. Phys. Rev. E **76**, 031102 (2007) 77
2. Beck, C.: Dynamical foundations of nonextensive statistical mechanics. Phys. Rev. Lett. **87**, 180601 (2001) 77
3. Beck, C., Cohen, E.: Superstatistics. Physica A **322**, 267–275 (2003) 77
4. Chavanis, P.H., Sire, C.: Statistics of velocity fluctuations arising from a random distribution of point vortices: The speed of fluctuations and the diffusion coefficient. Phys. Rev. E **62**, 490–506 (2000) 72
5. Crooks, G.: Beyond Boltzmann-Gibbs statistics: Maximum entropy hyperensembles out-of-equilibrium. Phys. Rev. E **75**, 041119 (2007) 77
6. Rajagopal, A.K.: Superstatistics - a quantum generalization. arXiv:cond-mat/0608679 (2006) 77
7. Van der Straeten, E., Beck, C.: Superstatistical distributions from a maximum entropy principle. Phys. Rev. E **78**, 051101 (2008) 77

Chapter 6
The Mean Field Approximation

6.1 The Ideal Paramagnet

Consider a number of spin variables $\sigma_1, \sigma_2, \cdots, \sigma_N$, which each can take the values ± 1. A one-parameter family of probability distributions is defined by

$$q_\epsilon(\sigma) = \prod_{n=1}^{N} \epsilon^{\frac{1}{2}(1+\sigma_n)}(1-\epsilon)^{\frac{1}{2}(1-\sigma_n)}, \qquad 0 < \epsilon < 1. \tag{6.1}$$

This probability distribution is of the *product* type. Each spin is independent of all others and has the value $+1$ with probability ϵ, the value -1 with probability $1-\epsilon$. In the mathematics literature such a probability distribution is often called *iid*, which stands for independent and identically distributed.

The distribution (6.1) belongs to the exponential family. Indeed, one can write

$$q_\epsilon(\sigma) = \exp\left(-\Phi(\theta) - \theta H_0(\sigma)\right) \tag{6.2}$$

with

$$\theta = \frac{1}{2h} \ln \frac{\epsilon}{1-\epsilon}, \tag{6.3}$$

$$\Phi(\theta) = -\frac{N}{2} \ln \epsilon(1-\epsilon) = \frac{N}{2} \ln 4 + N \ln \cosh(h\theta), \tag{6.4}$$

and

$$H_0(\sigma) = -hM \qquad \text{with } M = \sum_{n=1}^{N} \sigma_n. \tag{6.5}$$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_6,

The constant $h > 0$ is arbitrary. It has the meaning of an external magnetic field favouring the value $+1$ of the spin variables. To obtain the above relations we used that the inverse function of (6.3) is

$$\epsilon = \frac{1}{1+e^{-2h\theta}} \qquad \text{and} \qquad 1-\epsilon = \frac{1}{1+e^{2h\theta}}. \tag{6.6}$$

The average energy equals

$$\begin{aligned} U \equiv \langle H_0 \rangle &= -\frac{\partial \Phi}{\partial \theta} = -Nh\tanh(h\theta) \\ &= -Nh(2\epsilon - 1). \end{aligned} \tag{6.7}$$

The Boltzmann-Gibbs-Shannon entropy of (6.1) is

$$\begin{aligned} S(q_\epsilon) &= -\sum_\sigma q_\epsilon(\sigma)\ln q_\epsilon(\sigma) \\ &= -N\left[\epsilon\ln\epsilon + (1-\epsilon)\ln(1-\epsilon)\right]. \end{aligned} \tag{6.8}$$

The inverse temperature is given by

$$\begin{aligned} \beta &= \frac{\mathrm{d}S}{\mathrm{d}U} = \frac{\mathrm{d}S}{\mathrm{d}\epsilon}\frac{\mathrm{d}\epsilon}{\mathrm{d}U} \\ &= -\frac{1}{2h}\ln\frac{\epsilon}{1-\epsilon} \\ &= \frac{1}{2h}\ln\frac{Nh-U}{Nh+U}. \end{aligned} \tag{6.9}$$

This is positive when $U < 0$, which is the case if $1/2 < \epsilon < 1$, or equivalently, $\theta > 0$. [1] Inverting this relation gives

$$U = -Nh\tanh(h\beta). \tag{6.10}$$

Comparison with (6.7) shows that $\beta = \theta$. The heat capacity equals

$$C = \frac{\mathrm{d}U}{\mathrm{d}T} = N\beta^2 h^2(1 - \tanh^2\beta h)). \tag{6.11}$$

See the Figure 6.1. The positivity of the heat capacity ($C > 0$) confirms that the entropy $S(U)$ is a concave function of energy U, as it should be for thermodynamic *stability* reasons.

[1] Alternatively, one can change the definition of H_0 by taking $h < 0$ in (6.5). Then, $0 < \epsilon < 1/2$ corresponds with positive temperatures.

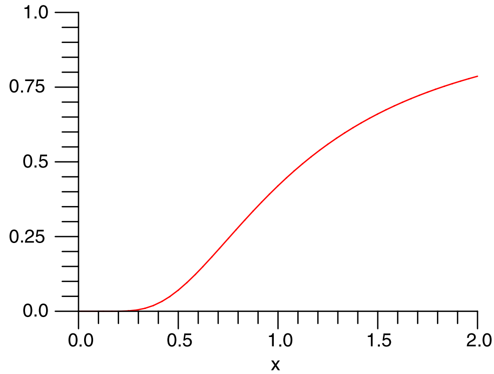

Fig. 6.1 Heat capacity of the ideal paramagnet as a function of temperature T. Plotted is $1 - \tanh^2(1/x)$ — see (6.11)

6.2 The Mean Field Equation

The energy of an interacting system of *Ising spins* is

$$H(\sigma) = -\frac{1}{2}\sum_{m,n=1}^{N} J_{m,n}\sigma_m\sigma_n - h\sum_{m=1}^{N}\sigma_m. \tag{6.12}$$

The first term of this expression is the interaction term. It is absent in the Hamiltonian $H_0(\sigma)$ of the ideal paramagnet — see (6.5).

The relevant thermodynamic quantities are now $U = \langle H\rangle$ and $M = \langle\sum_{m=1}^{N}\sigma_m\rangle$. However, let us keep the same family of probability distributions as in the previous section, this is, the distributions of the product form. The main argument to do so is that it is very convenient to work with product measures. The price one pays is that correlations between different spins are neglected. Indeed, such correlations are produced by the interaction term but are not present in the ideal paramagnet. In this situation, the thermodynamic variable $H(\sigma)$ is not the variable $H_0(\sigma)$ appearing in the canonical expression (6.2). Hence the stability results of the previous chapter are not any longer guaranteed. This point will be clarified later on.

Because the distributions q_ϵ are of the product form one has for $m \neq n$

$$\langle\sigma_m\sigma_n\rangle_\epsilon = \langle\sigma_m\rangle\langle\sigma_n\rangle_\epsilon = (2\epsilon - 1)^2. \tag{6.13}$$

Without restriction one can assume that $J_{m,m} = 0$. With this assumption the average energy equals

$$\langle H \rangle_\epsilon = -\frac{1}{2} N J (2\epsilon - 1)^2 - N h (2\epsilon - 1), \tag{6.14}$$

where $J \equiv (1/N) \sum_{m,n=1}^{N} J_{m,n}$. This result is used to maximise the expression

$$\begin{aligned} S(q_\epsilon) - \beta \langle H \rangle_\epsilon = & -N\epsilon \ln \epsilon - N(1-\epsilon) \ln(1-\epsilon) \\ & + \frac{1}{2} \beta N J (2\epsilon - 1)^2 + \beta h N (2\epsilon - 1), \end{aligned} \tag{6.15}$$

for given values of the control parameters β and h (that β is indeed the inverse temperature will be checked later on). In this way one looks for the optimal value of the parameter ϵ.

Variation of (6.15) with respect to ϵ gives

$$0 = -\ln \frac{\epsilon}{1-\epsilon} + 2\beta J (2\epsilon - 1) + 2\beta h. \tag{6.16}$$

This can be written as (using $\langle \sigma_n \rangle = 2\epsilon - 1$)

$$\langle \sigma_n \rangle = \tanh \beta \left(J \langle \sigma_n \rangle + h \right). \tag{6.17}$$

This is the well-known *mean field equation*. It is an implicit equation for the average value $\langle \sigma_n \rangle$ of the spin variable σ_n. The solutions of this equation are discussed below.

6.3 Phase Transitions

Note that in (6.15) the entropy functional $S(p)$ is *not* maximised over all probability distributions having the right average energy, but only over the model distributions q_ϵ, which are probability distributions of the product form. As a consequence, the optimal value of $S(q_\epsilon)$, which is assumed to coincide with the thermodynamic entropy $S(U)$, does not have the properties that have been proved in Chapter 3 for the solution of the variational principle. In particular, the relation between the partial derivatives $\partial S / \partial U_k$ and the model parameters ϵ^k does not necessarily hold. But the relation (3.17) is essential for thermodynamic stability to hold — see the discussion at the end of Section 3.7. It is therefore not a surprise that in mean field models a *thermodynamic instability* can occur. In the context of Chapter 3, this is, assuming a model belonging to the exponential family, thermodynamic stability is automatically satisfied. In that context phase transitions can only occur as limiting cases.

By definition, a *phase transition* of order n is a discontinuity in the n-th derivative of the free energy with respect to the thermodynamic variables. A well-known example is the phase transition in the two-dimensional Ising model in absence of an external magnetic field. It is not possible to write the free energy $F_N(T, h = 0)$ of the Ising model in a closed form for arbitrary size $N = L \times L$ of the square lattice. However, Kramers and Wannier [3] succeeded in showing that, in order to calculate the free energy per spin in the limit of large N

$$f(T) = \lim_{N \to \infty} \frac{1}{N} F_N(T, h = 0) \tag{6.18}$$

it suffices to calculate the largest eigenvalue of a matrix of dimensions $2^N \times 2^N$, the so-called *transfer matrix*. Subsequently, Lars Onsager [6] succeeded to calculate this largest eigenvalue. The result is that $f(T)$ is known in closed form. It turns out that it is a real analytic function on the interval $(0, T_c)$, as well as on the interval $(T_c, +\infty)$, with a singularity at the critical temperature T_c, determined by

$$\sinh(2\beta_c J) = 1. \tag{6.19}$$

Numerically, this gives $k_B T_c \simeq 2.2692J$. See for instance [10], Section 5.5.

6.4 A Mean Field Phase Transition

Let us return to the mean field model. The solution of the mean field equation (6.17) yields the average magnetisation $M(\beta, h) = N\langle\sigma_n\rangle$ as a function of inverse temperature β and external field h. For convenience, introduce the notation $m(\beta, h) = M(\beta, h)/N$. The Massieu function $\Phi(\beta, h)$ follows using (6.15). The result is

$$\frac{1}{N}\Phi(\beta, h) = -\frac{1+m}{2}\ln\frac{1+m}{2} - \frac{1-m}{2}\ln\frac{1-m}{2} + \frac{1}{2}\beta J m^2 + \beta h m. \tag{6.20}$$

First consider the situation without external field, i.e. $\beta h = 0$. Then the obvious solution of the mean field equation is $m = 0$. This gives

$$\frac{1}{N}\Phi(\beta, h = 0) = \ln 2, \tag{6.21}$$

independent of β. However, at low temperatures, this is, for β large enough, the mean field equation has two other solutions as well. The graphical solution of the mean field equation

$$m = \tanh(\beta J m), \qquad h = 0, \tag{6.22}$$

makes this clear. In Figure 6.2 the left and right hand sides are plotted as a function of m. The intersection points of the two curves are the solutions of

(6.22). One sees that there is only one intersection point when the slope of $\tanh(\beta J m)$ in the point $m = 0$ is small. There are three intersection points when the slope is larger than 1. The Taylor expansion of the tanh function is

$$\tanh(u) = u - \frac{1}{3}u^3 + \mathrm{O}(u^5). \tag{6.23}$$

Hence the transition from one to three solutions occurs when $\beta J = 1$. One concludes that the mean field model is able to predict the phase transition that actually occurs in the 2-dimensional Ising model. However, the prediction of the value of the phase transition temperature is way off. One can argue that the interaction strength J of the mean field model is not the same as that of the Ising model and should be corrected by a factor 2. But even then there remains a difference of about 10% between the mean field prediction $\beta_c J = 0.5$ and the result of the Ising model $\beta_c J \simeq 1/2.2692$.

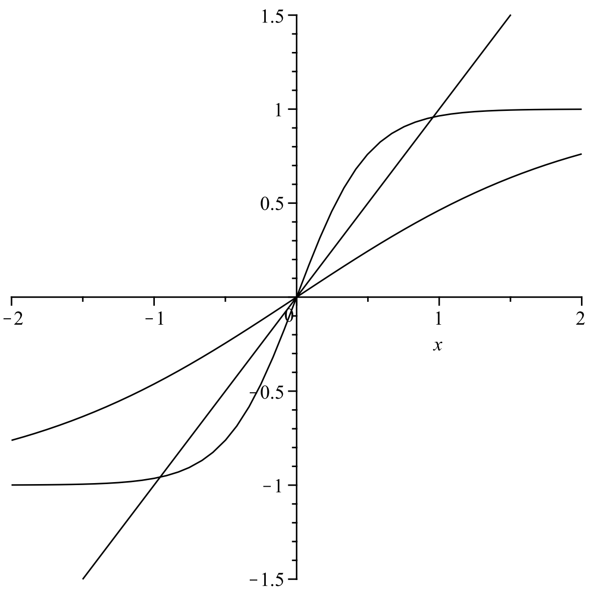

Fig. 6.2 Plot of x, of $\tanh(x/2)$, and of $\tanh(2x)$

Note that (6.14) can be written as

$$U = -\frac{1}{2N} J M^2 - hM \tag{6.24}$$

Hence, U and M are *not* independent thermodynamic variables, as it should be. This means that, by restricting the family of probability distributions to certain product measures, only one line in the two-dimensional thermodynamic configuration space is probed. Indeed, with a single parameter ϵ one can only probe a one-dimensional subset of this two-dimensional space. As a consequence, the dependence of entropy $S(U, M)$ on energy U, and separately on magnetisation M, cannot be determined in a unambiguous manner. However, one can calculate the Massieu function $\Phi(\beta, h)$ as a function of inverse temperature β and external field h (or, equivalently, the free energy $F(T, h) = -T\Phi(\beta, h)$ as a function of T and h). It is equal to the maximum value obtained in (6.15). Let $M(\beta, h)$ be the solution of the mean field equation (6.17) which maximises (6.15). By substituting $\epsilon = (1 + m)/2$ in (6.15) one obtains

$$\begin{aligned}\frac{1}{N}\Phi(\beta, h) = &-\frac{1}{2}(1+m)\ln\frac{1}{2}(1+m) - \frac{1}{2}(1-m)\ln\frac{1}{2}(1-m)\\ &+\frac{1}{2}\beta J m^2 + \beta h m.\end{aligned} \tag{6.25}$$

This function has a singularity in the point $\beta = \beta_c$ and $h = 0$. Indeed, for $0 < \beta < \beta_c$ and $h = 0$ one has $m(\beta, h) = 0$ so that $\Phi(\beta, h)/N$ is constant, equal to $\ln 2$. For $\beta > \beta_c$ and $h = 0$ it is not constant.

Let us evaluate this singularity. From the mean field equation (6.22) and the series expansion (6.23) there follows

$$m = \beta J m - \frac{1}{3}(\beta J m)^3 + \cdots. \tag{6.26}$$

This implies that either $m = 0$ or

$$m \simeq \pm\frac{\sqrt{3}}{\beta J^{3/2}}\sqrt{T_c - T}. \tag{6.27}$$

The latter solutions are real only when $T < T_c$. See the Figure 6.3.

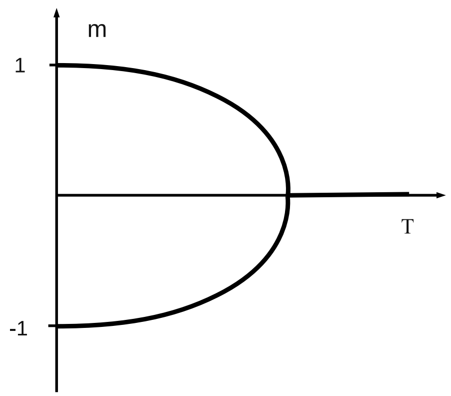

Fig. 6.3 Plot of the spontaneous magnetisation m as a function of the temperature T

It will become clear in the next Section that the solution $m = 0$ is not acceptable when $T < T_c$. Hence, the sudden change from $m = 0$ above the phase transition temperature to $m \neq 0$ below implies that the Massieu function (6.25) is not an analytic function in the point $\beta = \beta_c, h = 0$.

6.5 A Hyperensemble of Product States

Consider now a hyperensemble of the product states $q_\epsilon(\sigma)$, defined by (6.1). The entropy $\check{\mathcal{S}}(\epsilon)$ is taken equal to the Boltzmann-Gibbs entropy $S(p_\epsilon)$. The hyperdistribution $f(\epsilon)$, after optimisation, becomes – see (5.27) and (6.15)

$$\begin{aligned} f_{\beta,h}(\epsilon) &= \frac{1}{Z(\beta,h)} \exp\left(S(p_\epsilon) - \beta\langle H\rangle_\epsilon\right) \\ &= \frac{1}{Z(\beta,h)} \exp\bigg(- N\epsilon\ln\epsilon - N(1-\epsilon)\ln(1-\epsilon) \\ &\qquad + \frac{1}{2}\beta NJ(2\epsilon-1)^2 + \beta hN(2\epsilon-1)\bigg). \end{aligned} \tag{6.28}$$

The partition sum equals

$$\begin{aligned} Z(\beta,h) = \int_0^1 \mathrm{d}\epsilon\, \exp\bigg(&- N\epsilon\ln\epsilon - N(1-\epsilon)\ln(1-\epsilon) \\ &+ \frac{1}{2}\beta NJ(2\epsilon-1)^2 + \beta hN(2\epsilon-1)\bigg). \end{aligned} \tag{6.29}$$

In the limit of large N this integral can be evaluated because the only contribution comes from the value of ϵ for which the argument of the exponential function in (6.29) is maximal. But this search for a maximum is precisely what has been done in Section 6.2 and what leads to the mean field equation (6.17). The result is that $p_{\beta,h}(\sigma) \simeq q_\epsilon(\sigma)$, where ϵ has the value which optimises (6.15). This shows that for large systems the approach using hyperensembles reproduces the mean field approximation. In fact, for $h = 0$ and $\beta > \beta_{\mathrm{mf}}$ the hyperdistribution $f_{\beta,h}(\epsilon)$ has two maxima while for high temperatures it has a single maximum at $\epsilon = 1/2$. See the Figure 6.4.

6.6 Generalised Mean Field Theories

The mean field theory starts from an ensemble of product states q_η together with an entropy functional $\check{\mathcal{S}}(\eta)$, which is for example the Boltzmann entropy or its modification. However, the approach with hyperensembles allows for more general choices of q_η and $\check{\mathcal{S}}(\eta)$. To illustrate this point let us consider the two-state *Markov chain*. This looks very much like a one-dimensional Ising

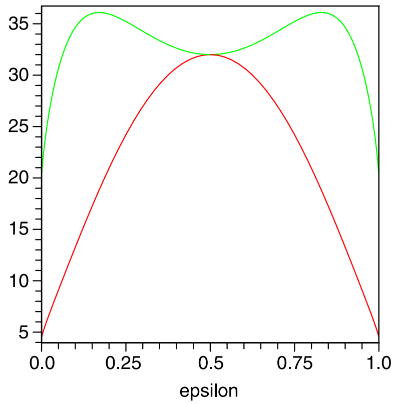

Fig. 6.4 Hyperdistribution $f_{\beta,h}(\epsilon)$ with $N = 5$ for $h = 0$, $\beta = 0.6$ and $\beta = 1.2$

chain. It consists of a stochastic variable σ and of transition probabilities ϵ and μ. The stochastic variable takes on two values $+1$ and -1. The transition probabilities determine the conditional probability that the stochastic variable σ_n at time n has the same value as at time $n-1$. See the Figure 6.5.

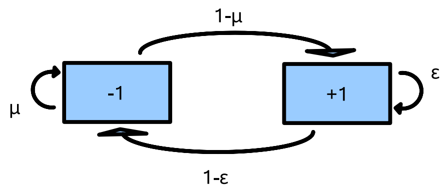

Fig. 6.5 State diagram of the two-state Markov chain

Consider now a path of length n of the Markov chain. This is a sequence of values ± 1 taken on by the stochastic variables $\sigma_0, \sigma_1, \cdots, \sigma_n$. For example, $++-+-+$ is a path of 5 steps, where a plus stands for $+1$, a minus for -1. The probability of such a path is the probability of the starting element times a polynomial of the form

$$\epsilon^{k_{++}}(1-\epsilon)^{k_{+-}}\mu^{k_{--}}(1-\mu)^{k_{-+}}, \tag{6.30}$$

where k_{rt} counts the number of times that a state with value r is followed by a state with value t.

The interesting quantities are now the total time spent in each of the two states and the number of transitions from one state to the other. The former is related to

$$M = \sum_{i=0}^{n} \sigma_i, \tag{6.31}$$

the latter to

$$H = -\sum_{i=1}^{n} \sigma_{i-1}\sigma_i. \tag{6.32}$$

Let $q_{\epsilon\mu}^{(\pm)}(x, E)$ denote the probability that $M - \sigma_0$ (neglecting the initial state) has the value x and H has the value E, given that the Markov chain starts in $\sigma_0 = s$. It can be written as

$$q_{\epsilon\mu}^{(s)}(x, E) = \sum_k C_s(k) e^{-\Psi(k)} \tag{6.33}$$

with

$$\Psi(k) = -k_{++} \ln \epsilon - k_{+-} \ln(1-\epsilon) - k_{--} \ln \mu - k_{-+} \ln(1-\mu), \tag{6.34}$$

where $C_s(k)$ is a counting factor that keeps track of the number of times that a certain k-vector occurs. Now note that

$$\begin{aligned} n &= k_{++} + k_{+-} + k_{-+} + k_{--} \\ x &= k_{++} - k_{+-} + k_{-+} - k_{--} \\ -E &= k_{+-} + k_{-+}. \end{aligned} \tag{6.35}$$

Introduce the variable $\Delta = k_{+-} - k_{-+}$. One has $\Delta = 0$ or $\Delta = \pm 1$. Then one can write

$$\begin{aligned} \Psi(k) = &-\frac{1}{2} n \ln(\epsilon\mu) - \frac{1}{2} x \ln \frac{\epsilon}{\mu} - \frac{1}{2} E \ln \frac{\epsilon\mu}{(1-\epsilon)(1-\mu)} \\ &-\frac{1}{2} \Delta \ln \frac{\epsilon(1-\epsilon)}{\mu(1-\mu)}. \end{aligned} \tag{6.36}$$

The results obtained so far have two drawbacks. The calculation of the counting factors $C_s(k)$ is possible [11] but rather cumbersome. In addition, the expression (6.36) still depends on k via the function $\Delta(k)$. These problems can be circumvented by adopting *stationary initial conditions*. These are given by $\text{Prob}(\sigma_0 = \pm 1) = p_\pm^{(0)}$ with

$$p_+^{(0)} = \frac{1-\mu}{2-\epsilon-\mu}, \quad \text{and} \quad p_-^{(0)} = \frac{1-\epsilon}{2-\epsilon-\mu}. \tag{6.37}$$

Let

$$q_{\epsilon\mu}(x, E) = p_+^{(0)} q_{\epsilon\mu}^{(+)}(x, E) + p_-^{(0)} q_{\epsilon\mu}^{(-)}(x, E). \tag{6.38}$$

Let $\langle\cdot\rangle_{\epsilon,\mu}$ denote the expectation value with respect to $q_{\epsilon\mu}(x,E)$. Then by symmetry one has $\langle\Delta\rangle_{\epsilon,\mu}=0$. By taking derivatives with respect to ϵ and μ of the identities $1=\sum_{x,E} q^{(s)}_{\epsilon\mu}(x,E)$, one obtains (see [12])

$$\begin{aligned}\langle x\rangle_{\epsilon\mu} &= n\frac{\epsilon-\mu}{2-\epsilon-\mu}\\ \langle E\rangle_{\epsilon\mu} &= -2n\frac{(1-\epsilon)(1-\mu)}{2-\epsilon-\mu}.\end{aligned} \tag{6.39}$$

This leads to

$$\begin{aligned}\langle\Psi\rangle_{\epsilon\mu} &= -\frac{1}{2}n\ln(\epsilon\mu)-\frac{1}{2}\langle x\rangle\ln\frac{\epsilon}{\mu}-\frac{1}{2}\langle E\rangle\ln\frac{\epsilon\mu}{(1-\epsilon)(1-\mu)}\\ &= np_+^{(0)}[-\epsilon\ln\epsilon-(1-\epsilon)\ln(1-\epsilon)]\\ &\quad +np_-^{(0)}[-\mu\ln\mu-(1-\mu)\ln(1-\mu)].\end{aligned} \tag{6.40}$$

This quantity is the *dynamical entropy* of the Markov chain — see [5]. Taking $\check{\mathcal{S}}=\langle\Psi\rangle_{\epsilon\mu}$, the hyperdistribution becomes (omitting the dependence on the parameters β and βF)

$$f(\epsilon,\mu)=\frac{1}{Z}\exp(\langle\Psi\rangle_{\epsilon\mu}-\beta\langle E\rangle_{\epsilon\mu}-\beta F\langle x\rangle_{\epsilon\mu}). \tag{6.41}$$

Extrema of this distribution occur when

$$\begin{aligned}\beta &= -\frac{1}{2}\ln\frac{\epsilon\mu}{(1-\epsilon)(1-\mu)}\\ \beta F &= -\frac{1}{2}\ln\frac{\epsilon}{\mu}.\end{aligned} \tag{6.42}$$

These equations have a unique solution satisfying $\epsilon>0$ and $\mu>0$.

6.7 The Quantum Case

The equivalent of product measures in the quantum case are tensor products of density operators. However, this works only for distinguishable particles. For example, the spins of atoms in a lattice can be distinguished by there positions. The electron exchange between neighbouring atoms can be modelled phenomenologically as an interaction between spins. This leads to the quantum Heisenberg model — see the Box 6.1. The mean field equations of this model reduces to the classical mean field equation, discussed above. The mean field treatment is in agreement with a hyperensemble based on the von Neumann entropy $\check{\mathcal{S}}(\eta)=-\operatorname{Tr}\sigma_\eta\ln\sigma_\eta$.

Consider quantum spins, labelled $1, 2, \cdots, N$, described by Pauli matrices $\sigma_1, \sigma_2, \sigma_3$. The three Pauli matrices of the n-th spin are placed together in a $3d$-vector $\overline{\sigma}_n$. The matrices $\sigma_{n\alpha}$ and $\sigma_{m\gamma}$ with $m \neq n$ commute with each other. With these notations, the Hamiltonian of the Heisenberg model is

$$H = -\frac{1}{2}\sum_{m,n=1}^{N} J_{m,n}\overline{\sigma}_m \cdot \overline{\sigma}_n - h\sum_{n=1}^{N}\sigma_{n3}. \tag{6.43}$$

Without restriction, assume $J_{nn} = 0$.

The single-spin density matrix

$$\rho_\theta = \frac{1}{2}\left(\mathbf{I} - \theta^k\sigma_k\right) \tag{6.44}$$

has been introduced before — see (2.9). Let $\rho_\theta^{\otimes N}$ denote the tensor product of N copies of ρ_θ. It is again a density matrix. One has $\mathrm{Tr}\,\rho_\theta^{\otimes N}\sigma_{n\alpha} = \mathrm{Tr}\,\rho_\theta\sigma_\alpha = -\theta^\alpha$ and, for $m \neq n$, $\mathrm{Tr}\,\rho_\theta^{\otimes N}\overline{\sigma}_m \cdot \overline{\sigma}_n = \sum_\alpha (\mathrm{Tr}\,\rho_\theta\overline{\sigma}_\alpha)^2 = |\theta|^2$. Therefore, the average energy equals

$$U = \mathrm{Tr}\,\rho_\theta^{\otimes N} H = -\frac{1}{2}NJ|\theta|^2 + hN\theta^3, \tag{6.45}$$

with $J = \sum_{m,n=1}^{N} J_{m,n}/N$. On the other hand, the von Neumann entropy of this density matrix equals

$$\begin{aligned} S(\rho_\theta^{\otimes N}) &= -N\,\mathrm{Tr}\,\rho_\theta \ln \rho_\theta \\ &= -\frac{N}{2}(1+|\theta|)\ln\frac{1}{2}(1+|\theta|) - \frac{N}{2}(1-|\theta|)\ln\frac{1}{2}(1-|\theta|) \end{aligned} \tag{6.46}$$

(to obtain this result, first diagonalise ρ_θ before calculating $\rho_\theta \ln \rho_\theta$). Now optimise $S(\rho_\theta^{\otimes N}) - \beta\,\mathrm{Tr}\,\rho_\theta^{\otimes N} H$. This leads to the mean field equations

$$\frac{1}{2}\frac{\theta^\alpha}{|\theta|}\ln\frac{1+|\theta|}{1-|\theta|} = -2\beta J\theta^\alpha - h\delta_{\alpha,3}. \tag{6.47}$$

These equations reduce to $\theta^1 = \theta^2 = 0$ and $\theta^3 = \tanh(\beta J\theta^3 + \beta h)$. The latter coincides with the classical result (6.17).

Box 6.1 The Heisenberg model

The situation is different for indistinguishable particles, either fermions or bosons. They are usually discussed in Fock space. The density operators that are easy to treat then correspond with *quasifree states*. These should replace the product states as starting point for the mean field approximation. However, it seems that such an approach has not been tried out.

Notes

A traditional exposition about the mean field approximation is found in [9], Chapter 6. However, the use of product measures as a way of doing mean field theory is known since long. The translation in terms of hyperensembles is of course new.

The treatment of the two-state Markov chain is taken from the papers [11, 12, 5, 13].

The notion of quasifree states is often used in the physics literature. However, they are usually not called quasifree. Rather they are associated with a Gaussian approximation. Sometimes they are called Gaussian states. In the boson case the quasifree states are related to the *coherent states*. The latter are usually defined either by a condition of minimal uncertainty or by wavefunctions that are eigenstates of the annihilation operator. The name of quasifree states comes out of the mathematical physics literature and was introduced in 1965 [8, 4, 1]. A mathematical introduction to quasifree states is found in volume II of Ola Bratteli and Derek Robinson [2], or, in the boson context, in the book of Dénes Petz [7].

Objectives

- Calculate the heat capacity of the ideal paramagnet.
- Derive the mean field equation for a ferromagnet.
- Know the definition of first and second order phase transitions.
- Discuss the graphical solution of the mean field equation for a ferromagnet.
- Know that Lars Onsager succeeded in finding an exact expression for the phase transition temperature of an Ising model with nearest-neighbour interactions on a square lattice.
- Show that the mean field approximation is a hyperensemble of an ensemble of product states.
- Know the definition of a two-state Markov chain.
- Apply the hyperensemble construction to other than ensembles of product states.
- Treat the quantum Heisenberg model in the mean field approximation.

References

1. Balslev, E., Verbeure, A.: States on Clifford algebras. Commun. Math. Phys. **7**, 55–76 (1968) 91
2. Bratteli, O., Robinson, D.: Operator Algebras and Quantum Statistical Mechanics. Vol I, II. 2nd edition. Springer, New York, Berlin (1997) 91

3. Kramers, H., Wannier, G.: Statistics of the two-dimensional ferromagnet. Part I, II. Phys. Rev. **60**, 252–276 (1941) 83
4. Manuceau, J., Verbeure, A.: Quasi-free states of the CCR-algebra and Bogoliubov transformations. Commun. Math. Phys. **9**, 293–302 (1968) 91
5. Naudts, J., Van der Straeten, E.: Transition records of stationary Markov chains. Phys. Rev. E **74**, 040103(R) (2006) 89, 91
6. Onsager, L.: Crystal statistics, I. A two-dimensional model with an order-disorder transition. Phys. Rev. **65**, 117–149 (1944) 83
7. Petz, D.: An invitation to the algebra of canonical commutation relations. Leuven University Press, Leuven (1990) 91
8. Robinson, D.: The ground state of the Bose gas. Commun. Math. Phys. **1**, 159–174 (1965) 91
9. Stanley, H.: Phase Transitions and Critical Phenomena. Harper & Row (1971) 91
10. Thompson, C.: Mathematical Statistical Mechanics. McMillan, New York (1972) 83
11. Van der Straeten, E., Naudts, J.: A one-dimensional model for theoretical analysis of single molecule experiments. J. Phys. A: Math. Gen. pp. 5715–5726 (2006) 88, 91
12. Van der Straeten, E., Naudts, J.: A two-parameter random walk with approximate exponential probability distribution. J. Phys. A: Math. Gen. **39**, 7245–7256 (2006) 89, 91
13. Van der Straeten, E., Naudts, J.: Residual entropy in a model for the unfolding of single polymer chains. Europhys. Lett. **81**, 28007 (2008) 91

Part II
Deformed Exponential Families

This part of the book deals with the notion of a generalised exponential family. It grew out of the research in the area of non-extensive thermostatistics. For that reason Tsallis' non-extensive thermostatistics is treated in the Section 8.

One of the conclusions is that the q-deformed exponential family occurs in a natural way within the context of classical mechanics. The more abstract generalisations discussed in the final chapters may seem less important from a physics point of view. But they have been helpful in elucidating the structure of the theory of generalised exponential families.

Chapter 7
q-Deformed Distributions

7.1 q-Deformed Exponential and Logarithmic Functions

An obvious way to generalise the Boltzmann-Gibbs distribution is by replacing the exponential function of the Gibbs factor by a function with similar properties. This function is then called a *deformed exponential*. The inverse of the deformed exponential function is the *deformed logarithmic function*. It turns out to be advantageous to start with deforming the logarithm.

Fix a number $q > 0$. The q-deformed logarithmic function is denoted $\ln_q(u)$. It is only defined for $u > 0$ and is the unique solution of the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}u} \ln_q(u) = \frac{1}{u^q} \tag{7.1}$$

which satisfies $\ln_q(1) = 0$. Note that this definition implies that the q-logarithm is an increasing function.

One recognises immediately for $q = 1$ the well-known property of the natural logarithm $(\mathrm{d}/\mathrm{d}u) \ln u = 1/u$. Hence, $\ln_1(u) \equiv \ln(u)$. For arbitrary $q \neq 1$ is

$$\begin{aligned} \ln_q(u) &= \int_1^u \mathrm{d}y\, y^{-q} \\ &= \frac{1}{1-q} \left(u^{1-q} - 1 \right). \end{aligned} \tag{7.2}$$

The second derivative of (7.1) is

$$\frac{\mathrm{d}^2}{\mathrm{d}u^2} \ln_q(u) = -q u^{-q-1}. \tag{7.3}$$

This is negative if $q > 0$ (remember that $u > 0$). A function with negative second derivative is *concave*, with positive second derivative it is *convex*. The

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_7,

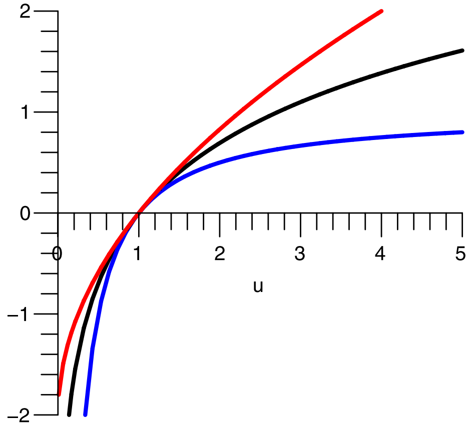

Fig. 7.1 The q-deformed logarithm $\ln_q(u)$ for q-values (from top to bottom) of 0.5, 1, and 2

natural logarithm $\ln(u)$ is concave. Since the q-deformed logarithm should look somewhat similar to the natural one, it is obvious to require that q is positive. However, sometimes negative q-values will be used.

Note that for large u and for $1 < q$ the q-deformed logarithm does not diverge but goes to $1/(q-1)$. Similarly, for small u and for $0 < q < 1$ it does not diverge to $-\infty$ but tends to $-1/(1-q)$. Hence, the natural logarithm is the only q-logarithm that diverges both for small and for large values of its argument. See the Figure 7.1. As a consequence, the inverse function, which is the q-deformed exponential function, and which is denoted $\exp_q(u)$, is not everywhere defined, except when $q = 1$. Since this is rather inconvenient, it is obvious to extend the definition of $\exp_q(u)$ to the whole of the real axis with a value of zero or $+\infty$, whatever is appropriate, in such a way that it is an increasing function on all of the real axis. More precisely, $\exp_q(u) = +\infty$ holds when $1 < q$ and $u > 1/(q-1)$; $\exp_q(u) = 0$ holds when $0 < q < 1$ and $u < -1/(1-q)$. By inverting (7.2) one obtains

$$\exp_q(u) = [1 + (1-q)u]_+^{1/(1-q)}. \tag{7.4}$$

The symbol $[\cdot]_+$ is the positive part and is defined by

$$[u]_+ = \max\{0, u\}. \tag{7.5}$$

In particular, the q-deformed exponential function is never negative. But it can be $+\infty$. The latter is indeed the case if the argument of $[\cdot]_+$ is negative or zero and the exponent $1/(q-1)$ is negative. See the Figure 7.2. With this

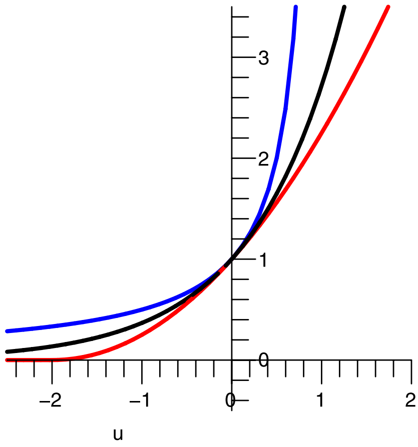

Fig. 7.2 The q-deformed exponential $\exp_q(u)$ for q-values (from top to bottom) 2.0, 1, and 0.5

extended definition of $\exp_q(u)$ one still has that for all u

$$\exp_q(\ln_q(u)) = u \tag{7.6}$$

However, the relation

$$\ln_q(\exp_q(u)) = u \tag{7.7}$$

is only defined when $\exp_q(u)$ is neither 0 or $+\infty$.

The basic properties of the q-deformed exponential, needed further on, are that $\exp_q(0) = 1$ and that it is an increasing and convex function. Note further that

$$\frac{\mathrm{d}}{\mathrm{d}u} \exp_q(u) = \left[\exp_q(u)\right]^q . \tag{7.8}$$

7.2 Dual Definitions

The well-known relations

$$\ln \frac{1}{u} = -\ln u \quad \text{and} \quad \exp(-u) = \frac{1}{\exp(u)} \tag{7.9}$$

are in general not valid for deformed exponential and logarithmic functions. Introduce therefore dual functions

$$\exp_q^*(u) = \frac{1}{\exp_q(-u)} \tag{7.10}$$

and

$$\ln_q^*(u) = -\ln_q(1/u). \tag{7.11}$$

A short calculation shows that $\exp_q^*(u) = \exp_{2-q}(u)$ and $\ln_q^*(u) = \ln_{2-q}(u)$. These are again deformed exponential and logarithmic functions provided that $0 < q < 2$.

The $q \leftrightarrow 2-q$-duality plays an important role in what follows. It should not be confused with the $q \leftrightarrow 1/q$-duality which is related to the notion of escort probabilities, discussed later on in the present Chapter.

7.3 The q-Exponential Family

The obvious definition for a family of probability distributions p_θ to belong to the *q-exponential family* is that it can be written into the form

$$p_\theta(x) = c(x)\exp_q\left(-\alpha(\theta) - \theta^k H_k(x)\right). \tag{7.12}$$

Like in the $q = 1$-case, it is essential that the prior weight $c(x)$ and the Hamiltonians $H_k(x)$ do not depend on the parameters θ, while the normalisation $\alpha(\theta)$ must not depend on the variable x. We write $\alpha(\theta)$ instead of $\Phi(\theta)$ because the normalisation does not necessarily coincide with the Massieu function. The parameter q is the *deformation index* of the family.

Using the explicit definition of the q-deformed exponential one can write (7.12) as

$$\begin{aligned} p_\theta(x) &= c(x)\left[1-(1-q)(\alpha(\theta)+\theta^k H_k(x))\right]_+^{1/(1-q)} \\ &= \frac{c(x)}{Z(\eta)}\left[1-(1-q)\eta^k H_k(x)\right]_+^{1/(1-q)} \\ &= \frac{c(x)}{Z(\eta)}\exp_q\left(-\eta^k H_k(x)\right), \end{aligned} \tag{7.13}$$

with

$$\eta^k = \frac{\theta^k}{1-(1-q)\alpha(\theta)} = \theta^k Z(\eta)^{1-q} \tag{7.14}$$

and

Here we prove the statement that a probability distribution $p_\theta(x)$ belonging to the q-exponential family is a distribution of superstatistics if and only if $q \geq 1$ and (7.17) is satisfied.

The function $\tilde{f}(z)$ must clearly be given by

$$\tilde{f}(z) = \exp_q(-\alpha(\beta) - \beta z) = [1 + (q-1)(\alpha(\beta) + \beta z)]_+^{1/(1-q)} . \tag{7.18}$$

The first derivative of this function is

$$\frac{\mathrm{d}}{\mathrm{d}z}\tilde{f}(z) = -\left(\alpha(\beta) + \beta z\right)[1 + (q-1)(\alpha(\beta) + \beta z)]_+^{q/(1-q)} . \tag{7.19}$$

This must be strictly negative if $\tilde{f}(z)$ is completely monotonic. Hence, the requirement (7.17) is necessary.

A short calculation shows that the prefactor of the n-th derivative is

$$q(1-2q)(2-3q)\cdots(n-2-(n-1)q). \tag{7.20}$$

If $q > 1$ then this factor has alternating sign. However, if $0 < q < 1$ then there exists n such that $n - 2 - (n-1)q > 0$. When this is the case, then the n-th derivative has the same sign as the $n-1$-th. This is impossible if $\tilde{f}(z)$ is completely monotonic. Therefore, $q > 1$ is needed.

Box 7.1 Superstatistics and q-exponential families

$$\begin{aligned} Z(\eta) &= \frac{1}{\exp_q(-\alpha(\theta))} = \exp_q^*(\alpha(\theta)) \\ &= \int \mathrm{d}x\, c(x) \exp_q\left(-\eta^k H_k(x)\right) . \end{aligned} \tag{7.15}$$

Both formulas (7.12) and (7.13) are useful. The former expression implies

$$\ln_q(p_\theta(x)/c(x)) = -\alpha(\theta) - \theta^k H_k(x), \tag{7.16}$$

which will be used below. On the other hand, (7.13) allows for an easy calculation of the normalisation $Z(\eta)$. Once $Z(\eta)$ is known, the θ^k follow from (7.14).

A probability distribution $p_\beta(x)$, which belongs to the q-exponential family with one parameter β, and with Hamiltonian $H(x)$, is a distribution of *superstatistics*, this means, it is of the form (5.23), with a completely monotonic function $\tilde{f}(z)$, if and only if $q \geq 1$ and

$$\alpha(\beta) + \beta H(x) > 0 \qquad \text{for all } x. \tag{7.17}$$

See the Box 7.1. Note that for $q > 1$ the condition (7.17) is usually satisfied because Hamiltonians of physics are assumed to be bounded from below.

7.4 Escort Probabilities

Let us try to express that $p_\theta(x)$ is normalised to 1 for all choices of θ^k. This will lead us to the notion of escort probabilities.

From (7.16) and the definition of the q-logarithm follows

$$-\frac{\partial\alpha}{\partial\theta^k} - H_k(x) = \frac{\partial}{\partial\theta^k}\ln_q\frac{p_\theta(x)}{c(x)}$$
$$= \left(\frac{p_\theta(x)}{c(x)}\right)^{-q}\frac{1}{c(x)}\frac{\partial}{\partial\theta^k}p_\theta(x). \tag{7.21}$$

Hence, one finds

$$0 = \frac{\partial}{\partial\theta^k}\int \mathrm{d}x\, p_\theta(x)$$
$$= \int \mathrm{d}x\,\frac{\partial}{\partial\theta^k}p_\theta(x)$$
$$= \int \mathrm{d}x\, c(x)\left(\frac{p_\theta(x)}{c(x)}\right)^q\left(-\frac{\partial\alpha}{\partial\theta^k} - H_k(x)\right)$$
$$= -z(\theta)\frac{\partial\alpha}{\partial\theta^k} - \int \mathrm{d}x\, c(x)\left(\frac{p_\theta(x)}{c(x)}\right)^q H_k(x), \tag{7.22}$$

with

$$z(\theta) = \int \mathrm{d}x\, c(x)\left(\frac{p_\theta(x)}{c(x)}\right)^q. \tag{7.23}$$

Introduce now a new parametrised family of probability distributions, denoted $P_\theta(x)$, and defined by

$$P_\theta(x) = \frac{c(x)}{z(\theta)}\left(\frac{p_\theta(x)}{c(x)}\right)^q. \tag{7.24}$$

Averages with respect to this new family will be denoted with double brackets

$$\langle\langle f\rangle\rangle_\theta = \int \mathrm{d}x\, P_\theta(x) f(x). \tag{7.25}$$

This new family is called the *escort probability distribution.*

Now, one can write (7.22) as

$$\frac{\partial\alpha}{\partial\theta^k} = -\langle\langle H_k\rangle\rangle_\theta. \tag{7.26}$$

This shows that the variables H_k still can be used as (biased) estimators to estimate the parameters θ. However, in evaluating the empirical average of the H_k one should use the escort probabilities instead of the original p_θ.

Expression (7.26) should be compared with (2.19), which involves the averages $\langle H_k \rangle_\theta$ with respect to the p_θ instead of the escort family and which holds for the Massieu function $\Phi(\theta)$ instead of the normalisation $\alpha(\theta)$. Only when $q = 1$ both functions coincide.

The explicit expression of the escort probability family is

$$P_\theta(x) = \frac{c(x)}{z(\theta)} \left[1 - (1-q)(\alpha(\theta) + \theta^k H_k(x))\right]_+^{q/(1-q)} . \tag{7.27}$$

Note that it produces quantities relevant for comparison with experimental data! Indeed, the average energies $\langle\langle H_k \rangle\rangle_\theta$ are calculated using the escort probabilities and are used to estimate the parameters via (7.26).

7.5 q-to-1/q Duality and Escort Families

The definition (7.24) of the escort probability can be inverted to give

$$p_\theta(x) = c(z) z(\theta)^{1/q} \left(\frac{P_\theta(x)}{c(x)}\right)^{1/q} . \tag{7.28}$$

This suggests that the escort of $P_\theta(x)$ is again $p_\theta(x)$. To make this meaningful, $P_\theta(x)$ should belong to the deformed exponential family with deformation index $1/q$ instead of q. Note that one can write

$$\begin{aligned} P_\theta(x) &= \frac{c(x)}{z(\theta)} \left[1 + (1-q)(-\alpha(\theta) - \theta^k H_k(x)\right]_+^{q/(1-q)} \\ &= \frac{c(x)}{z(\theta)} \frac{1}{\exp_{1/q}\left[q\alpha(\theta) + q\theta^k H_k(x)\right]} . \end{aligned} \tag{7.29}$$

This equals $\exp_{1/q}(-\alpha(\theta) - \theta^k H_k(x))$ only if $q = 1$, in which case $P_\theta(x)$ and $p_\theta(x)$ coincide. Hence, in general, $P_\theta(x)$ does not belong to the $1/q$-exponential family.

If $q > \frac{1}{2}$ then the dual exponential function $\exp_q^*(u) = \exp_{2-q}(u)$ is a deformed exponential. This makes it possible to write

$$\begin{aligned} P_\theta(x) &= \frac{c(x)}{z(\theta)} \exp_{1/q}^*\left[-q\alpha(\theta) - q\theta^k H_k(x)\right] \\ &= \frac{c(x)}{z(\theta)} \exp_{2-1/q}\left[-q\alpha(\theta) - q\theta^k H_k(x)\right] . \end{aligned} \tag{7.30}$$

In this case $P_\theta(x)$ belongs to the deformed exponential family with deformation index $2 - 1/q$. Its escort probability is therefore proportional to $P_\theta(x)^{2-1/q}$ and hence to $p_\theta(x)^{2q-1}$. Only for $q = 1$ is the latter equal to $p_\theta(x)$.

One concludes that replacing $p_\theta(x)$ by its escort $P_\theta(x)$ is not a symmetry of the deformed exponential families. Rather, $p_\theta(x)$ and its escort $P_\theta(x)$ should be seen as two equivalent descriptions of the same model. The probability distribution $p_\theta(x)$ belongs to the q-deformed exponential family if and only if its escort $P_\theta(x)$ belongs to the *escort family* with deformation index $1/q$. The latter is defined by the fact that after reparametrisation it can be written into the form

$$P_\zeta(x) = c(x)\frac{1}{\exp_{1/q}\left(\alpha^*(\zeta) + \zeta^k H_k(x)\right)}. \tag{7.31}$$

The relation with (7.29) is then

$$\begin{aligned} \zeta^k &= qz(\theta)\theta^k \\ \alpha^*(\zeta) &= \frac{q}{q-1}(z(\theta) - 1) + qz(\theta)\alpha(\theta). \end{aligned} \tag{7.32}$$

7.6 Dual Identities

A nice property of a probability distribution belonging to the exponential family is that the average energies $\langle H_k \rangle_\theta$ can be calculated by taking partial derivatives of the logarithm of the partition sum — see (2.19) in Chapter 2. This property can be generalised to distributions belonging to the q-exponential family and will be generalised in Chapter 11 to an even larger class of models. However, as noted above, the partial derivatives of the normalisation function $\alpha(\theta)$, although the latter plays the role of the logarithm of the partition sum, do not return the average energies but rather the averages with respect to the escort probability — see (7.26). This raises the question whether a function $\Phi(\theta)$ exists such that (2.19) holds.

Let us start with the observation that the pdfs $p_\theta(x)$ of the q-exponential family maximise the expression

$$I(p) - s\theta^k \langle H_k \rangle_p, \tag{7.33}$$

with $s = 2 - q$, where the entropy functional $I(p)$ is given by

$$\begin{aligned} I(p) &= -\frac{1}{1-q}\int \mathrm{d}x\, p(x)\left[\left(\frac{p(x)}{c(x)}\right)^{1-q} - 1\right] \\ &= -\int \mathrm{d}x\, p(x) \ln_q\left(\frac{p(x)}{c(x)}\right). \end{aligned} \tag{7.34}$$

The proof of this statement is given in the more general context of Chapter 11 — see Section 11.5. The constant s is introduced to allow that entropy functionals are defined up to a constant positive factor. The alternative of forcing $s = 1$ in (7.33) requires that $I(f)$ should be divided by $2 - q$. But when $q > 2$ this means multiplying the entropy functional with a negative constant, which is not desirable.

It is now obvious to define the entropy $S(U)$ by

$$S(U) = I(p_\theta) \quad \text{where} \quad U_k = \langle H_k \rangle_\theta. \tag{7.35}$$

A short calculation gives

$$S(U) = I(p_\theta) = \alpha(\theta) + \theta^k U_k. \tag{7.36}$$

The Legendre transform of $S(U)$ is the function $\Phi(\theta)$. It is given by

$$\begin{aligned}\Phi(\theta) &= S(U) - s\theta^k U_k \\ &= \alpha(\theta) - (1-q)\theta^k U_k.\end{aligned} \tag{7.37}$$

Because it is a Legendre transform one has automatically the validity of the relation

$$\frac{\partial \Phi}{\partial \theta^k} = -sU_k, \tag{7.38}$$

and of the dual relation

$$\frac{\partial S}{\partial U_k} = s\theta^k. \tag{7.39}$$

In the standard case $q = 1$ Massieu's function equals the logarithm of the partition sum. Hence, to know the expected values U_k it suffices to calculate the partition sum. When $q \neq 1$ the expression (7.37) contains a correction term. Hence, in general it is not enough to know the normalisation $\alpha(\theta)$ to obtain expressions for the U_k. Indeed, as we have seen in (7.26), the derivatives of the normalisation $\alpha(\theta)$ yield averages w.r.t. the escort probability distribution instead of the $p_\theta(x)$.

7.7 The q-Gaussian Distribution

The *q-Gaussian distribution* in one variable is given by

$$f(x) = \frac{1}{c_q \sigma} \exp_q(-x^2/\sigma^2), \tag{7.40}$$

with

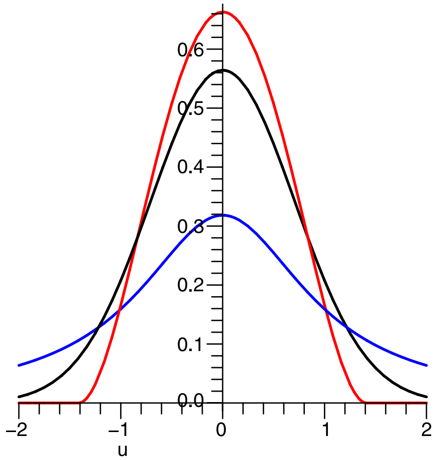

Fig. 7.3 q-Gaussians for $\sigma = 1$ and for q-values 0.5, 1, and 2

$$c_q = \int_{-\infty}^{\infty} \mathrm{d}x \, \exp_q(-x^2) = \sqrt{\frac{\pi}{q-1}} \frac{\Gamma\left(-\frac{1}{2} + \frac{1}{q-1}\right)}{\Gamma\left(\frac{1}{q-1}\right)} \quad \text{if} \quad 1 < q < 3,$$
$$= \sqrt{\frac{\pi}{1-q}} \frac{\Gamma\left(1 + \frac{1}{1-q}\right)}{\Gamma\left(\frac{3}{2} + \frac{1}{1-q}\right)} \quad \text{if} \quad q < 1. \tag{7.41}$$

It belongs to the q-exponential family. Indeed, it can be brought into the form (7.12) with $c(x) = 1/c_q$, $H(x) = x^2$, $\theta = \sigma^{q-3}$, and

$$\alpha(\theta) = \frac{\sigma^{q-1} - 1}{q-1} = \ln_{2-q}(\sigma). \tag{7.42}$$

The $q = 1$-case reproduces the conventional Gauss distribution. For $q < 1$ the distribution vanishes outside an interval. Take for instance $q = 1/2$. Then (7.40) becomes

$$f(x) = \frac{15\sqrt{2}}{32\sigma} \left[1 - \frac{x^2}{\sigma^2}\right]_+^2. \tag{7.43}$$

This distribution vanishes outside the interval $[-\sigma, \sigma]$. In the range $1 \le q < 3$ the q-Gaussian is strictly positive on the whole line and decays with a power law in $|x|$ instead of exponentially. For $q = 2$ one obtains

$$f(x) = \frac{1}{\pi} \frac{\sigma}{x^2 + \sigma^2}. \tag{7.44}$$

This is the Cauchy distribution. The function (7.44) is also called a Lorentzian and is often used in physics to fit the shape of spectral lines. For $q \geq 3$ the distribution cannot be normalized because

$$f(x) \sim \frac{1}{|x|^{2/(q-1)}} \text{ as } |x| \to \infty \tag{7.45}$$

is not integrable any more.

The q-Maxwellian

The most important example of q-Gaussians in statistical physics is the velocity distribution in a classical gas with N particles. It is a *q-Maxwellian* which only in the limit of large N converges to the *Maxwell distribution* (1.1).

The Hamiltonian of an N-particle classical ideal gas is given by

$$H(p) = \frac{1}{2m} \sum_{j=1}^{N} |p_j|^2, \tag{7.46}$$

where m is the mass of the particles and p_j is the momentum of the j-th particle. Given the value E for the total energy, the *microcanonical* phase space consists of all points on the surface of a $3N$-dimensional sphere with radius $\sqrt{2mE}$. Let $B_n(r)$ denote the volume of a sphere with radius r in dimension n. The probability distribution for the momentum of a single particle becomes

$$\begin{aligned} f(p_1) &= \frac{1}{(2m)^{3N/2} B'_{3N}(\sqrt{E})} \int \mathrm{d}^3 p_2 \cdots \mathrm{d}^3 p_N \, \delta(E - H(p)) \\ &= \frac{1}{(2m)^{3/2}} \frac{B'_{3(N-1)}(\sqrt{E - |p_1|^2/2m})}{B'_{3N}(\sqrt{E})} \\ &= \frac{A}{(2\sigma^2)^{3/2}} \exp_q(-|p_1|^2/2\sigma^2) \end{aligned} \tag{7.47}$$

with normalisation constant A and with

$$q = \frac{3N-6}{3N-4} \quad \text{and} \quad \sigma^2 = \frac{2mE}{3N-4}. \tag{7.48}$$

This is a q-Gaussian with $q < 1$. The appearance of a cutoff can be easily understood. Arbitrary large momenta are not possible because of the obvious upperbound $|p_1|^2 \leq 2mE$. The probability distribution of the scalar velocity $v = |p_1|/m$ becomes

$$f_q(v) = \frac{1}{d_q}\left(\frac{m}{\sigma^2}\right)^{3/2} v^2 \exp_q\left(-\frac{1}{2\sigma^2}mv^2\right), \tag{7.49}$$

with $d_q = \int_0^\infty \mathrm{d}u\, u^2 \exp_q(-u/2)$. Only in the limit of large systems this distribution converges to the Maxwell distribution. One concludes that the Maxwell distribution is an approximation and is only valid in the limit of large systems ($N \to \infty$).

7.8 The configurational probability distribution of a mono-atomic gas

The proof, given above, that in a finite and isolated gas of particles the probability distribution of the scalar velocities is a q-Maxwellian rather than a Maxwellian, can be generalised to show that in the microcanonical ensemble the configurational probability distribution of any *mono-atomic gas* belongs to the q-exponential family.

The microcanonical ensemble is described by the singular probability density function (see Chapter 4)

$$q_E(\mathbf{q},\mathbf{p}) = \frac{1}{\omega(E)}\delta(E - H(\mathbf{q},\mathbf{p})), \tag{7.50}$$

where $\delta(\cdot)$ is Dirac's delta function. The normalization is so that

$$1 = \frac{1}{N!h^{3N}}\int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{p}_1\cdots\mathrm{d}\mathbf{p}_N \int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{q}_1\cdots\mathrm{d}\mathbf{q}_N\, q_E(\mathbf{q},\mathbf{p}). \tag{7.51}$$

For simplicity, we take only one conserved quantity into account, namely the total energy. Its value is fixed to E. The *density of states* equals

$$\omega(E) = \frac{1}{N!h^{3N}}\int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{p}_1\cdots\mathrm{d}\mathbf{p}_N \int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{q}_1\cdots\mathrm{d}\mathbf{q}_N\, \delta(E - H(\mathbf{q},\mathbf{p})). \tag{7.52}$$

It is in principle possible to integrate out the momenta. This leads to the *configurational probability distribution*, which is given by

$$q_E^{\mathrm{conf}}(\mathbf{q}) = \left(\frac{a}{h}\right)^{3N}\int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{p}_1\cdots\mathrm{d}\mathbf{p}_N\, q_E(\mathbf{q},\mathbf{p}). \tag{7.53}$$

The normalization is so that

$$1 = \frac{1}{N!a^{3N}}\int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{q}_1\cdots\mathrm{d}\mathbf{q}_N q_E^{\mathrm{conf}}(\mathbf{q}). \tag{7.54}$$

The constant a has been introduced for dimensional reasons.

In the simplest case the Hamiltonian is of the form

$$H(\mathbf{q},\mathbf{p}) = \frac{1}{2m}\sum_{j=1}^{N}|\mathbf{p}_j|^2 + \mathcal{V}(\mathbf{q}), \tag{7.55}$$

where $\mathcal{V}(\mathbf{q})$ is the potential energy due to interaction among the particles and between the particles and the walls of the system. Then the integration over the momenta can be carried through explicitly. One obtains

$$\begin{aligned} q_E^{\mathrm{conf}}(\mathbf{q}) &= \left(\frac{a}{h}\right)^{3N}\frac{1}{\omega(E)}\frac{\mathrm{d}}{\mathrm{d}E}\int_{\mathbf{R}^{3N}}\mathrm{d}\mathbf{p}_1\cdots\mathrm{d}\mathbf{p}_N\,\Theta\left(E-\mathcal{V}(\mathbf{q})-\frac{1}{2m}\sum_{j=1}^{N}|\mathbf{p}_j|^2\right)\\ &= \left(\frac{a}{h}\right)^{3N}\frac{1}{\omega(E)}\frac{\mathrm{d}}{\mathrm{d}E}B_{3N}\left(\sqrt{2m(E-\mathcal{V}(\mathbf{q}))}\right), \end{aligned} \tag{7.56}$$

where $B_n(r)$ is the volume of the sphere with radius r in dimension n. Because $B_n(r) = r^n B_n(1)$, one can continue with

$$\begin{aligned} q_E^{\mathrm{conf}}(\mathbf{q}) &= \left(\frac{a}{h}\right)^{3N}\frac{1}{\omega(E)}\frac{\mathrm{d}}{\mathrm{d}E}B_{3N}(1)\left[2m(E-\mathcal{V}(\mathbf{q}))\right]^{3N/2}\\ &= \left(\frac{a}{h}\right)^{3N}\frac{1}{\omega(E)}B_{3N}(1)(2m)^{3N/2}\frac{3N}{2}\left[E-\mathcal{V}(\mathbf{q})\right]^{3N/2-1}. \end{aligned} \tag{7.57}$$

In order to write this in the form of a member of the q-exponential family one has to bring the E-dependent factor $1/\omega(E)$ inside the expression to the power $3N/2-1$. In addition, let

$$q = 1 - \frac{1}{\frac{3N}{2}-1} = \frac{3N-4}{3N-2}. \tag{7.58}$$

Then the above result becomes

$$q_E^{\mathrm{conf}}(\mathbf{q}) = \left(\frac{a}{h}\right)^{3N}B_{3N}(1)(2m)^{3N/2}\frac{3N}{2}\left[\omega(E)^{q-1}\left\{E-\mathcal{V}(\mathbf{q})\right\}\right]^{\frac{1}{1-q}}. \tag{7.59}$$

Introduce the parameter

$$\theta = \frac{1}{1-q}\left(\frac{2ma^2}{h^2}\right)^{2-q}\omega(E)^{q-1}. \tag{7.60}$$

The extra factor in front of $\omega(E)^{q-1}$ has been chosen so that θ is an inverse energy and that (7.59) can now been written as a q-exponential with a dimensionless argument. Indeed, one has now

$$q_E^{\mathrm{conf}}(\mathbf{q}) = \frac{3N}{2}B_{3N}(1)\exp_q(-\alpha(E)-\theta\mathcal{V}(\mathbf{q})) \tag{7.61}$$

with

$$\alpha(E) = \frac{1}{1-q}\left[1 - \left(\frac{2ma^2}{h^2}\right)^{2-q}\frac{E}{\omega(E)^{1-q}}\right]. \tag{7.62}$$

It is now clear that the configurational probability distribution $q_E^{\mathrm{conf}}(\mathbf{q})$ belongs to the q-exponential family.

7.9 Average Kinetic Energy

Because the configurational density function belongs to the q-exponential family, it satisfies the dual identities (7.38, 7.38). In the present example they reproduce the statement that the ratio $\Omega(E)/\omega(E)$ of the integrated density of states too the density of states equals the average kinetic energy — see relation (4.15).

The entropy $I(p_\theta)$ evaluates to — see (7.36),

$$\begin{aligned} S^{\mathrm{conf}}(U^{\mathrm{conf}}) \equiv I(p_\theta) &= \alpha(\theta) + \theta U^{\mathrm{conf}} \\ &= \frac{1}{1-q} - \theta U^{\mathrm{kin}}, \end{aligned} \tag{7.63}$$

with $U^{\mathrm{kin}} = E - U^{\mathrm{conf}}$. The corresponding Massieu function is then given by

$$\begin{aligned} \Phi(\theta) &= S^{\mathrm{conf}}(U^{\mathrm{conf}}) - s\theta U^{\mathrm{conf}} \\ &= \frac{1}{1-q} - \theta U^{\mathrm{kin}} - s\theta U^{\mathrm{conf}}. \end{aligned} \tag{7.64}$$

Using the dual identities (7.38, 7.39) one obtains

$$sU^{\mathrm{conf}} = -\frac{\partial\Phi}{\partial\theta} = U^{\mathrm{kin}} + sU^{\mathrm{conf}} + (1-q)\theta\frac{\mathrm{d}U^{\mathrm{conf}}}{\mathrm{d}\theta} + \theta\frac{\mathrm{d}E}{\mathrm{d}\theta} \tag{7.65}$$

and

$$s\theta = \frac{\mathrm{d}S^{\mathrm{conf}}}{\mathrm{d}U^{\mathrm{conf}}} = \theta - \left[\theta + U^{\mathrm{kin}}\frac{\mathrm{d}\theta}{\mathrm{d}E}\right]\frac{\mathrm{d}E}{\mathrm{d}U^{\mathrm{conf}}}. \tag{7.66}$$

Since both identities imply the same result we continue with one of them. Using

$$\frac{1}{\theta}\frac{\mathrm{d}\theta}{\mathrm{d}E} = (q-1)\frac{\omega'(E)}{\omega(E)} \tag{7.67}$$

the latter can be written as

$$\begin{aligned}\frac{\mathrm{d}U^{\mathrm{conf}}}{\mathrm{d}E} &= -\frac{1}{1-q} - \frac{1}{1-q}U^{\mathrm{kin}}\frac{1}{\theta}\frac{\mathrm{d}\theta}{\mathrm{d}E}\\ &= -\frac{1}{1-q} + \frac{\omega'(E)}{\omega(E)}U^{\mathrm{kin}}.\end{aligned} \tag{7.68}$$

Use this result to calculate

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}E}\omega(E)U^{\mathrm{kin}} &= \omega'(E)U^{\mathrm{kin}} + \omega(E)\left[1 - \frac{\mathrm{d}U^{\mathrm{conf}}}{\mathrm{d}E}\right]\\ &= \frac{2-q}{1-q}\omega(E) = \frac{3N}{2}\omega(E).\end{aligned} \tag{7.69}$$

By integrating this expression one obtains the average kinetic energy

$$U^{\mathrm{kin}} = \frac{3N}{2}\frac{\Omega(E)}{\omega(E)}. \tag{7.70}$$

This expression gives the relation between the average kinetic energy and the total energy E via the density of states $\omega(E)$ and its integral $\Omega(E)$. The integration constant must be taken so that $\Omega(E) = 0$ when $E = E_{\min}$ (implying that the kinetic energy vanishes in the ground state).

7.10 The Quantum Family

The quantum model with density operators ρ_θ belongs to the *quantum q-exponential family* if there exist self-adjoint operators H_k such that

$$\rho_\theta = \exp_q(-\alpha(\theta) - \theta^k H_k). \tag{7.71}$$

The function $\alpha(\theta)$ is used for normalisation. It is possible to take the normalisation in front of the deformed exponential. One has

$$\begin{aligned}\rho_\theta &= \exp_q\left(-\alpha(\theta) - \theta^k H_k\right)\\ &= \left[1 - (1-q)\alpha(\theta) - (1-q)\theta^k H_k\right]_+^{1/(1-q)}\\ &= \frac{1}{Z(\eta)}\left[1 - (1-q)\eta^k H_k\right]_+^{1/(1-q)}\\ &= \frac{1}{Z(\eta)}\exp_q(-\eta^k H_k),\end{aligned} \tag{7.72}$$

with η^k as in Section 7.3 on the quantum exponential family, equations (7.14, 7.15), and with $Z(\eta)$ given by

$$Z(\eta) = \mathrm{Tr}\,\exp_q(-\eta^k H_k). \tag{7.73}$$

It is obvious to define *escort density operators σ_q* by

$$\sigma_q = \frac{1}{z(\theta)} \rho_\theta^q, \tag{7.74}$$

with

$$z(\theta) = \operatorname{Tr} \rho_\theta^q. \tag{7.75}$$

(Assume that $\operatorname{Tr} \rho_\theta^q < +\infty$, which is not necessarily the case). As in the Section 7.5 on the $q \leftrightarrow 1/q$-duality, one can wonder whether the escort density matrix σ_q belongs to some q-deformed family. However, this need not always be the case. We do not repeat this discussion here.

The basic property of the escort density operators is their relation with the θ^k-derivatives of ρ_θ. However, as before, in Section 2.7, the operators H_k do not necessarily commute between each other. Hence, it is not straightforward to calculate the derivatives of ρ_θ. But using cyclic permutation under the trace one calculates

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta^k} \operatorname{Tr} \rho_\theta \\ &= \operatorname{Tr} \frac{\partial}{\partial \theta^k} \exp_q(-\alpha(\theta) - \theta^k H_k) \\ &= \operatorname{Tr} \rho_\theta^q \frac{\partial}{\partial \theta^k} \left[-\alpha(\theta) - \theta^l H_l\right] \\ &= z(\theta) \left[-\frac{\partial \alpha}{\partial \theta^k} - \langle\langle H_k \rangle\rangle_\theta \right], \end{aligned} \tag{7.76}$$

with

$$\langle\langle H_k \rangle\rangle_\theta = \operatorname{Tr} \sigma_\theta H_k. \tag{7.77}$$

This implies

$$-\frac{\partial \alpha}{\partial \theta^k} = \langle\langle H_k \rangle\rangle_\theta. \tag{7.78}$$

Hence, the quantum expectation of the estimators H_k with respect to the escort density operator can be used to estimate the parameters θ^k.

Problems

7.1. The kappa-distribution
The following distribution is known in plasma physics as the *kappa-distribution* (see for instance [5]). It is also called the *generalised Lorentzian distribution.*

$$f_\kappa(v) = \frac{1}{A(\kappa)v_0} \frac{c(v)}{\left(1 + \frac{1}{\kappa - a}\frac{v^2}{v_0^2}\right)^{1+\kappa}}, \quad \kappa > a. \tag{7.79}$$

Usually is $c(v) = v^2/v_0^2$. Write the kappa-distribution in the form of the q-Maxwellian (7.49), with $q = \dfrac{2+\kappa}{1+\kappa}$ and $\sigma^2 = \dfrac{1}{2}\dfrac{\kappa - a}{1+\kappa} m v_0^2$.

7.2. The Student's t-distribution
The density function of the *Student's t-distribution* is usually written as

$$f_n(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}. \tag{7.80}$$

The parameter n must be positive, not necessarily integer, and is called the *number of degrees of freedom*. (Often one writes the Greek letter ν instead of the Latin n. Here this would lead to confusion with the letter v used for the velocity).

The distribution arises in a natural manner when variables with normal fluctuations are added together. Show that the distribution $g_\lambda(v) = \lambda f(\lambda v)$ is a kappa-distribution with $\kappa = (1+n)/2$, $c(v) = 1$, $v_0 = \sqrt{2}/\lambda$, and $a = 1/2$ (see the previous Problem).

7.3. Order statistics
Select n numbers $u_1, u_2, \cdots, u_n$, uniformly chosen from the interval $[0,(n-1)T]$. Show that the probability distribution $p_T(u)$ of $u = \min\{u_1, u_2, \cdots, u_n\}$ belongs to the curved q-exponential family with $q = (n-2)/(n-1) < 1$ and $H(u) = u$.

Similarly, select n positive numbers according to the distributions

$$f(u) = \frac{c}{(1+cu)^2} \quad \text{with } c = \frac{1}{(n+1)T}. \tag{7.81}$$

Then $p_T(u)$ belongs to the curved q-exponential family with $q = (n+2)/(n+1) > 1$ and $H(u) = u$.

These examples are taken from [13] and belong to the domain of Order Statistics [4].

7.4. Stationary solutions
Let be given a potential $V(x)$ on the real line. Show that the probability distribution

$$p(x) = \exp_q(-\alpha(\theta) - \theta V(x)) \tag{7.82}$$

is a stationary solution of the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\frac{\partial V}{\partial x} p + \frac{\partial^2}{\partial x^2}(\lambda + \mu V)p. \tag{7.83}$$

Determine λ and μ. See [2] for the relation with multiplicative noise.

7.5. The ground state wave function of the relativistic harmonic oscillator reads

$$\psi_t(x) = Ae^{-i\hbar^{-1}(E_0 - mc^2)t}\left(1 + \frac{\omega^2 x^2}{c^2}\right)^{-(1/4)-mc^2/2\hbar\omega}, \tag{7.84}$$

with normalisation constant A and with

$$E_0 = \frac{1}{2}\hbar\omega + \sqrt{\frac{1}{4}(\hbar\omega)^2 + (mc^2)^2}. \tag{7.85}$$

Show that the spatial dependence of $|\psi_t(x)|^2$ is a q-Gaussian with $q > 1$ given by

$$\frac{1}{q-1} = \frac{1}{2} + \frac{mc^2}{\hbar\omega}. \tag{7.86}$$

See [1, 11].

7.6. Second moment of the q-Gaussian
a) Show that the second moment of the escort probability of the q-Gaussian is given by

$$\langle\langle x\rangle\rangle_\theta = \frac{1}{3-q}\sigma^2. \tag{7.87}$$

Use (7.26) to obtain this result.
b)Show that the second moment of the q-Gaussian itself satisfies the equation

$$(1-q)\theta\frac{\mathrm{d}U}{\mathrm{d}\theta} = U - \frac{1}{3-q}\theta^{-2/(3-q)} \quad \text{with } U = \langle x^2\rangle_\theta \text{ and } \theta = \sigma^{q-3} \tag{7.88}$$

To find this result, make use of the identity (7.38).

Notes

The idea of q-deformed logarithmic and exponential functions was first proposed by Constantino Tsallis [9]. A more systematic treatment, introducing the notion of dual deformed functions, is found in [6]. They are at the origin of the $q \leftrightarrow 2-q$ duality. The q-exponential family, together with a further generalisation, which is studied in Chapter 10, were introduced in [7]. The $q \leftrightarrow 1/q$ duality was discussed in [10]. The same paper also introduced the escort probability distributions into nonextensive statistical physics. This concept of escort measures was borrowed from a book on fractals [3], where however it plays a different role.

The q-Gaussians have been discussed by many authors. Christophe Vignat has pointed out [12] that q-Gaussians appear when, starting from a uniform distribution on a hypersphere, some of the variables are integrated out. This is essentially the reason why the q-exponential family appears in the microcanonical ensemble. That the configurational density can be written as a q-exponential is known since long. The remark that it belongs to the q-exponential family is found in [8].

Objectives

- Know the definitions of q-deformed logarithms and exponentials; understand the problem of defining $\exp_q(u)$ when u is not in the range of $\ln_q$; understand the notion of dual deformed functions.
- Know the definition of the q-exponential family, both classically and quantum mechanically.
- Why are escort probabilities needed?
- What is at the origin of the $q \leftrightarrow 1/q$ and the $q \leftrightarrow 2-q$ dualities?
- The probability distributions of the q-exponential family are distributions of superstatistics only when $q > 1$.
- Know that the q-exponential family appears in a natural way in the context of the classical microcanonical ensemble.

References

1. Aldaya, V., Bisquert, J., Guerrero, J., Navarro-Salas, J.: Group theoretic construction of the quantum relativistic harmonic oscillator. Rep. Math. Phys. **37**, 387–418 (1996) 112
2. Anteneodo, C., Tsallis, C.: Multiplicative noise: A mechanism leading to nonextensive statistical mechanics. J. Math. Phys. **44**, 5194 (2003) 112
3. Beck, C., Schlögl, F.: Thermodynamics of chaotic systems: an introduction. Cambridge University Press (1997) 113, 147
4. David, H.A., Nagaraja, H.: Order statistics. Wiley (2003) 111
5. Meyer-Vernet, N., Moncuquet, M., Hoang, S.: Temperature inversion in the Io-plasma torus. Icarus **116**, 202–213 (1995) 110
6. Naudts, J.: Deformed exponentials and logarithms in generalized thermostatistics. Physica A **316**, 323–334 (2002) 112, 162
7. Naudts, J.: Estimators, escort probabilities, and phi-exponential families in statistical physics. J. Ineq. Pure Appl. Math. **5**, 102 (2004) 112, 162, 177
8. Naudts, J., Baeten, M.: Non-extensivity of the configurational density distribution in the classical microcanonical ensemble. Entropy **11**, 285–294 (2009) 113
9. Tsallis, C.: What are the numbers that experiments provide? Quimica Nova **17**, 468 (1994) 112

10. Tsallis, C., Mendes, R., Plastino, A.: The role of constraints within generalized nonextensive statistics. Physica A **261**, 543–554 (1998) 112, 117, 129
11. Vignat, C., Lamberti, P.: A study of the orthogonal polynomials associated with the quantum harmonic oscillator on constant curvature spaces. J. Math. Phys. **50**, 103514 (2009) 112
12. Vignat, C., Plastino, A.: The p-sphere and the geometric substratum of power-law probability distributions. Phys. Lett. A **343**, 411–416 (2005) 113
13. Wilk, G., Włodarczyk, Z.: Tsallis distribution from minimally selected order statistics. In: S. Abe, H. Herrmann, P. Quarati, A. Rapisarda, C. Tsallis (eds.) Complexity, metastability and nonextensivity, *AIP Conference Proceedings*, vol. 965, pp. 76–79. American Institute of Physics (2007) 111

Chapter 8
Tsallis' Thermostatistics

8.1 The Tsallis Entropy Functional

The *Tsallis entropy functional* is defined by

$$\begin{aligned} S_q^{\mathrm{Tsallis}}(p) &= -\int \mathrm{d}x\, p(x) \ln_q^* \frac{p(x)}{c(x)} \\ &= \int \mathrm{d}x\, p(x) \ln_q \frac{c(x)}{p(x)} \\ &= \frac{1}{1-q}\left[\int \mathrm{d}x\, c(x)\left(\frac{p(x)}{c(x)}\right)^q - 1\right]. \end{aligned} \tag{8.1}$$

(See the definition of the dual function $\ln_q^*(u)$ in Section 7.2).

In the discrete case, without a prior weight $c(x)$, this reads

$$\begin{aligned} S_q^{\mathrm{Tsallis}}(p) &= \sum_j p(j) \ln_q \frac{1}{p(j)} \\ &= \frac{1}{1-q}\left(\sum_j p(j)^q - 1\right). \end{aligned} \tag{8.2}$$

This entropy functional was introduced in the Physics literature by Constantino Tsallis in 1988 [15], but was mentioned before by mathematicians [6, 5].

Note that

$$\begin{aligned} S_{2-q}^{\mathrm{Tsallis}}(p) &= \frac{1}{q-1}\left(\sum_j p(j)^{2-q} - 1\right) \\ &= -\sum_j p(j) \ln_q(p(j)). \end{aligned} \tag{8.3}$$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_8,

This expression coincides with (7.34).

In the $q = 1$-limit $S_q^{\text{Tsallis}}(p)$ reduces to the Boltzmann-Gibbs-Shannon entropy functional. To see this, use the theorem of de l'Hôpital

$$\begin{aligned}\lim_{q\to 1}\frac{1}{1-q}\left(\sum_j p(j)^q - 1\right) &= \lim_{q\to 1}\frac{\sum_j p(j)^q \ln p(j)}{-1}\\ &= -\sum_j p(j)\ln p(j).\end{aligned} \tag{8.4}$$

The concavity of the Tsallis entropy functional follows from the explicit expression of (8.1). The function which maps $p(x)/c(x)$ onto $(p(x)/c(x))^q$ is convex if $q > 1$ and concave if $0 < q < 1$. But the sign of the prefactor $1/(1-q)$ is negative in the former case and positive in the latter. Hence, the total expression is always concave.

If $q > 1$ then the q-logarithm has an upper bound $\ln_q(u) \le 1/(q-1)$. This implies an upper bound for the Tsallis entropy

$$S_q^{\text{Tsallis}}(p) \le \frac{1}{q-1}, \qquad q > 1. \tag{8.5}$$

If $0 < q < 1$ then $\ln_q(u) \ge -1/(1-q)$ for all $u > 0$. This implies

$$S_q^{\text{Tsallis}}(p) \ge -\frac{1}{1-q}, \qquad 0 < q < 1. \tag{8.6}$$

Hence, in this case the Tsallis entropy functional is bounded from below. However, this lower bound can be improved in the discrete case, in absence of prior weights $c(j)$. Indeed, for $0 < q < 1$ is

$$\begin{aligned}S_q^{\text{Tsallis}}(p) &= \frac{1}{1-q}\left(\sum_j p(j)^q - 1\right)\\ &\ge \frac{1}{1-q}\left(\sum_j p(j) - 1\right)\\ &= 0.\end{aligned} \tag{8.7}$$

Hence, in the case $0 < q < 1$ the discrete Tsallis entropy functional cannot become negative. To obtain this result, explicit use is made of the fact that the probabilities $p(j)$ cannot be larger than 1. In the continuous case the density $p(x)/c(x)$ can take on any positive value. For this reason, the argument working in the discrete case cannot be used in the continuous case.

A special case of the Tsallis entropy functional corresponds with $q = 2$, and is known as the *linear entropy*. It can be written as $S_2^{\text{Tsallis}}(p) = \sum_j p(j)(1 - p(j))$.

The Tsallis entropy functional is *non-additive* when $q \neq 1$. The Boltzmann-Gibbs-Shannon entropy is additive. This means that the entropy of a system composed of two independent subsystems A and B is the sum of the entropies of the two subsystems. More precisely, let $p(x, y) = p_{\mathrm{A}}(x)p_{\mathrm{B}}(y)$. Then one has $S(p) = S(p_{\mathrm{A}}) + S(p_{\mathrm{B}})$ The name of *non-extensive thermostatistics*, given to the domain of research to which the Tsallis entropy belongs, refers to the situation that the Boltzmann-Gibbs-entropy of a system does not grow linearly with the size of the system.

8.2 A Historical Reflection

In the early works on non-extensive thermostatistics the existence of dual definitions of deformed exponentials and logarithms was not noticed. As a consequence, it was not clear that the Tsallis entropy (8.1) is not naturally associated with the q-exponential family introduced in the previous Chapter. The natural candidate is (use that the dual logarithm satisfies $\ln_q^*(u) = \ln_{2-q}(u)$ to see this)

$$\begin{aligned} S_{2-q}^{\mathrm{Tsallis}}(p) &= \int \mathrm{d}x\, p(x) \ln_q^* \frac{c(x)}{p(x)} \\ &= -\int \mathrm{d}x\, p(x) \ln_q \frac{p(x)}{c(x)}. \end{aligned} \tag{8.8}$$

Indeed, using the definition (7.12), one finds immediately, see (7.36),

$$\begin{aligned} S_{2-q}^{\mathrm{Tsallis}}(p_\theta) &= -\int \mathrm{d}x\, p_\theta(x) \left(-\alpha(\theta) - \theta^k H_k(x)\right) \\ &= \alpha(\theta) + \theta^k \langle H_k \rangle_\theta. \end{aligned} \tag{8.9}$$

This expression looks like the result of a Legendre transform of $\alpha(\theta)$ and therefore suggests that the thermodynamic entropy $S(U)$, with energies U_k taken equal to $\langle H_k \rangle_\theta$, equals $S_{2-q}^{\mathrm{Tsallis}}(p_\theta)$. However, this cannot be correct because this would imply that the derivatives of $\alpha(\theta)$ equal $-\langle H_k \rangle_\theta$. But we know that they are equal to the escort averages $-\langle\langle H_k \rangle\rangle_\theta$ — see (7.26). The solution to this paradox is the constant $s = 2 - q$, introduced in the Section 7.6. The correct relation between the Massieu function $\Phi(\theta)$ and the entropy $S(U) = S_{2-q}^{\mathrm{Tsallis}}(p_\theta)$ is $S = \Phi + s\theta^k \langle H_k \rangle_\theta$. As a consequence, the normalisation $\alpha(\theta)$ is given by

$$\alpha(\theta) = \Phi(\theta) + (1 - q)\theta^k \langle H_k \rangle_\theta. \tag{8.10}$$

One sees immediately that $\alpha(\theta)$ and $\Phi(\theta)$ coincide in the limit $q = 1$.

In *non-extensive thermostatistics,* as formulated in the fundamental article [17], the entropy $S_q^{\mathrm{Tsallis}}(p_\theta)$ is associated with the escort energy $\langle\langle H_k \rangle\rangle_\theta$.

More precisely, the Tsallis entropy functional $S_q^{\text{Tsallis}}(p)$ is maximised under the constraint that the average of a given Hamiltonian $H(x)$ with respect to the escort probability $P(x)$ has some prescribed value U. The result of this rather peculiar procedure is a probability distribution, called the *Tsallis distribution*. This distribution belongs to the curved q-exponential family. The fact that it turns out to be curved explains why there has been quite some confusion in the literature about the correct definition of the thermodynamic temperature in Tsallis' thermostatistics. Indeed, in a curved q-exponential family the connection between the parameter θ and the inverse temperature β is further complicated by this curvature.

8.3 Maximising the Tsallis Entropy Functional

For simplicity, discrete probabilities are used in the present section and the prior weights $c(j)$ are omitted.

Consider the problem of maximising the Tsallis entropy functional $S_q^{\text{Tsallis}}(p)$ under the constraint that the escort probabilities

$$\langle\langle H_k\rangle\rangle_p \equiv \frac{1}{z_q(p)}\sum_j p(j)^q H_k(j) \quad \text{with } z_q(p) = \sum_j p(j)^q, \tag{8.11}$$

have given values U_k. This problem is solved by introducing Lagrange parameters θ^k and α, the latter to control the normalisation. The functional to be optimised is then

$$\mathcal{L}(p) = S_q^{\text{Tsallis}}(p) - \alpha\sum_j p(j) - \theta^k\langle\langle H_k\rangle\rangle_p. \tag{8.12}$$

Variation with respect to $p(j)$ gives

$$0 = \frac{q}{1-q}p_\theta(j)^{q-1} - \alpha - q\frac{p_\theta(j)^{q-1}}{z_q(\theta)}\theta^k\left[H_k(j) - \langle\langle H_k\rangle\rangle_\theta\right], \tag{8.13}$$

with

$$z_q(\theta) = \sum_j p_\theta(j)^q. \tag{8.14}$$

This expression can be written as

$$p_\theta(j) = \frac{1}{Z_q(\theta)}\left[1 - (1-q)\theta^k\frac{H_k(j) - \langle\langle H_k\rangle\rangle_\theta}{z_q(\theta)}\right]^{1/(1-q)}, \tag{8.15}$$

with the partition sum given by

$$Z_q(\theta) = \left(\alpha \frac{1-q}{q}\right)^{1/(1-q)}. \tag{8.16}$$

Here, we have silently assumed that the argument between square brackets is not negative so that it can be raised to the power $1/(1-q)$. This result can also be written in terms of a q-exponential

$$p_\theta(j) = \frac{1}{Z_q(\theta)} \exp_q \left[-\theta^k \frac{H_k(j) - \langle\langle H_k \rangle\rangle_\theta}{z_q(\theta)}\right]. \tag{8.17}$$

This probability distribution is the *Tsallis distribution*.

The partition sums $Z_q(\theta)$ and $z_q(\theta)$ are linked by an identity. Multiply (8.13) with $p_\theta(j)$ to obtain

$$0 = \frac{q}{1-q} p_\theta(j)^q - \alpha p_\theta(j) - q \frac{p_\theta(j)^q}{z_q(\theta)} \theta^k \left[H_k(j) - \langle\langle H_k \rangle\rangle_\theta\right]. \tag{8.18}$$

This can be written as

$$p_\theta(j) = \left(Z_q(\theta)\right)^{q-1} P_\theta(j) \left[z_q(\theta) - (1-q)\theta^k (H_k(j) - \langle\langle H_k \rangle\rangle_\theta)\right]. \tag{8.19}$$

Summing over j, and using that both the $p_\theta(j)$ and the escort $P_\theta(j)$ are properly normalised, yields the identity

$$z_q(\theta) = Z_q(\theta)^{1-q} = \alpha \frac{1-q}{q}. \tag{8.20}$$

Then, (8.19) simplifies to

$$p_\theta(j) = P_\theta(j) \left[1 - (1-q)\theta^k \frac{H_k(j) - \langle\langle H_k \rangle\rangle_\theta}{z_q(\theta)}\right]. \tag{8.21}$$

A nice property of the Tsallis distribution is that it is manifestly invariant under shifts of the energy scales: Adding a constant to the definition of any of the Hamiltonians $H_k(x)$ does not alter the probability distribution because only the difference between actual value and average value $\langle\langle H_k \rangle\rangle_\theta$ enters the expression.

8.4 Thermodynamic Properties

Note that (8.17) gives an implicit definition of the probabilities $p_\theta(j)$ because both sides of the equation depend on them. Nevertheless, the solution, when unique, defines a parametrised family of probability distributions. It belongs to the curved q-exponential family. This means that it can be written into the form

Let

$$\zeta^k(\theta) = \frac{Z_q(\theta)^{q-1}}{z_q(\theta)}\theta^k = \frac{1}{z_q(\theta)^2}\theta^k, \tag{8.23}$$

and

$$\alpha_0(\theta) = \frac{1 - Z_q(\theta)^{q-1}}{1-q} - \zeta^k(\theta)\langle\langle H_k\rangle\rangle_\theta. \tag{8.24}$$

Then (8.22) reduces to (8.17).

From the conservation of probability then follows

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta^k}\sum_j p_\theta(j) \\ &= \sum_j \frac{\partial}{\partial\theta^k}\exp_q\left(-\alpha_0(\theta) - \zeta^k(\theta)H_k(j)\right) \\ &= -\sum_j p_\theta(j)^q \frac{\partial}{\partial\theta^k}\left[\alpha_0(\theta) + \zeta^l(\theta)H_l(j)\right] \\ &= z(\theta)\left[-\frac{\partial\alpha_0}{\partial\theta^k} - \frac{\partial\zeta^l}{\partial\theta^k}\langle\langle H_l\rangle\rangle_\theta\right]. \end{aligned} \tag{8.25}$$

This can be written as

$$\frac{\partial\alpha_0}{\partial\theta^k} = -\frac{\partial\zeta^l}{\partial\theta^k}\langle\langle H_l\rangle\rangle_\theta, \tag{8.26}$$

and shows that the escort averages of the Hamiltonians H_k are not obtained simply by taking derivatives of the normalisation function with respect to the parameters θ. One needs in addition the matrix of coefficients $\partial\zeta^l/\partial\theta^k$.

Box 8.1 Proof that the Tsallis distribution belongs to the curved q-exponential family

$$p_\theta(j) = \exp_q\left(-\alpha_0(\theta) - \zeta^k(\theta)H_k(j)\right) \tag{8.22}$$

with $\alpha_0(\theta)$ and $\zeta^k(\theta)$ some functions of θ. See the Box 8.1.

Note further that there exists a simple relation between the value of the Tsallis entropy functional at equilibrium and the partition sum $z(\theta)$. Indeed, one calculates

$$\begin{aligned} S_q^{\text{Tsallis}}(p_\theta) &= \frac{1}{1-q}\left[\int \mathrm{d}x\, c(x)\left(\frac{p_\theta(x)}{c(x)}\right)^q - 1\right] \\ &= \frac{1}{1-q}\left[\int \mathrm{d}x\, z(\theta)P_\theta(x) - 1\right] \\ &= \frac{1}{1-q}\left[z(\theta) - 1\right]. \end{aligned} \tag{8.27}$$

Consider now a one-parameter model. Then the previous identity implies that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}S_q^{\mathrm{Tsallis}}(p_\theta) = \frac{1}{1-q}\frac{\mathrm{d}z}{\mathrm{d}\theta}. \tag{8.28}$$

On the other hand, one can show that

$$\theta\frac{\mathrm{d}U}{\mathrm{d}\theta} = \frac{1}{1-q}\frac{\mathrm{d}z}{\mathrm{d}\theta}. \tag{8.29}$$

Indeed, from (8.24) and (8.26) follows

$$\zeta\frac{\mathrm{d}U}{\mathrm{d}\theta} = \frac{1}{1-q}\frac{1}{z(\theta)^2}\frac{\mathrm{d}z}{\mathrm{d}\theta}. \tag{8.30}$$

Using (8.23) this simplifies to (8.29).

Identify now $S_q^{\mathrm{Tsallis}}(p_\theta)$ with the thermodynamic entropy $S(U)$. Combining both expression then yields (putting $k_B = 1$)

$$\beta \equiv \frac{\mathrm{d}S}{\mathrm{d}U} = \frac{\frac{\mathrm{d}}{\mathrm{d}\theta}S_q^{\mathrm{Tsallis}}(p_\theta)}{\frac{\mathrm{d}U}{\mathrm{d}\theta}} = \theta. \tag{8.31}$$

This proves that the parameter θ coincides with the inverse temperature β (assuming that $S_q^{\mathrm{Tsallis}}(p_\theta)$ coincides with the thermodynamic entropy $S(U)$).

8.5 Example: The Two-Level Atom

Consider a model with two states, described by a Hamiltonian with values $H(0) = 0$ and $H(1) = \Delta > 0$. This model is so simple that the step of maximising the entropy is unnecessary because the probability distribution is already fixed by the constraint. Indeed, from the constraint

$$U = \langle\langle H\rangle\rangle_\theta = P_\theta(1)\Delta \tag{8.32}$$

follows $P_\theta(1) = U/\Delta$. From $P_\theta(j) = p_\theta(j)^q/z(\theta)$ then follows

$$p_\theta(0) = \frac{\left(1-\frac{U}{\Delta}\right)^{1/q}}{\left(1-\frac{U}{\Delta}\right)^{1/q} + \left(\frac{U}{\Delta}\right)^{1/q}} \tag{8.33}$$

$$p_\theta(1) = \frac{\left(\frac{U}{\Delta}\right)^{1/q}}{\left(1-\frac{U}{\Delta}\right)^{1/q} + \left(\frac{U}{\Delta}\right)^{1/q}} \tag{8.34}$$

$$z(\theta) = \frac{1}{\left[\left(1-\frac{U}{\Delta}\right)^{1/q} + \left(\frac{U}{\Delta}\right)^{1/q}\right]^q}. \tag{8.35}$$

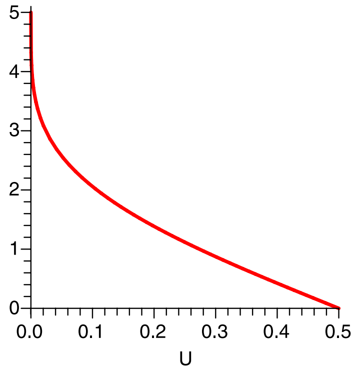

Fig. 8.1 Inverse temperature β as a function of the escort energy U, for $q = 0.8$ and $\Delta = 1$

Hence, the probability distribution and its escort are known as a function of U. The Tsallis entropy of this probability distribution equals

$$\begin{aligned} S_q^{\text{Tsallis}}(p_\theta) &= \frac{1}{1-q}\left[p_\theta(0)^q + p_\theta(1)^q - 1\right] \\ &= \frac{1}{1-q}\left[z(\theta) - 1\right]. \end{aligned} \tag{8.36}$$

Taking the derivative of the thermodynamic entropy $S(U)$ yields the inverse temperature β (taking $k_B = 1$). Identifying $S(U)$ with $S_q^{\text{Tsallis}}(p_\theta)$ then gives

$$\beta\Delta = \frac{\mathrm{d}S}{\mathrm{d}U}\Delta = \frac{1}{1-q}\frac{\left(1-\frac{U}{\Delta}\right)^{1/q-1} - \left(\frac{U}{\Delta}\right)^{1/q-1}}{\left[\left(1-\frac{U}{\Delta}\right)^{1/q} + \left(\frac{U}{\Delta}\right)^{1/q}\right]^{q+1}}. \tag{8.37}$$

This expression gives β as a function of U. See the Figure 8.1

Using the above expressions it is now a straightforward but tedious calculation to verify that p_θ can indeed be written as a Tsallis distribution with $\theta = \beta$. Hence, the example seems satisfactory. However, for smaller values of q, for instance $q = 0.5$, the inverse temperature β is not any longer a monotonic function of the escort energy U. This means that the heat capacity $\mathrm{d}U/\mathrm{d}T$ is negative and that the model is thermodynamically unstable. This is totally unexpected for such a simple model. It indicates that it is not correct to identify the escort energy U with the thermodynamic energy, and the Tsallis entropy $S_q^{\text{Tsallis}}(p_\theta)$ with the thermodynamic entropy $S(U)$, as we have done up to now. This point is further investigated in what follows.

8.6 Relative Entropy of the Csiszár Type

The f-*divergence*, or Csiszár type *divergence*, is defined for an arbitrary convex function $f(u)$ by

$$I(p||r) = \sum_k r(k) f\left(\frac{p(k)}{r(k)}\right). \tag{8.38}$$

The choice $f(u) = u \ln u$ yields

$$I(p||r) = \sum_k p(k) \ln\left(\frac{p(k)}{r(k)}\right). \tag{8.39}$$

This quantity is the *Kullback-Leibler distance*, see Section 3.6. In statistical physics, it is called the *relative entropy* of the distribution p with respect to the distribution r.

Make the choice

$$f(u) = \frac{u}{1-q}(1 - u^{q-1}) = -u \ln_q\left(\frac{1}{u}\right). \tag{8.40}$$

This function is convex for $q > 0$. It gives

$$I(p||r) = \frac{1}{1-q} \sum_k p(k) \left[1 - \left(\frac{r(k)}{p(k)}\right)^{1-q}\right]. \tag{8.41}$$

In the continuous case this becomes

$$I(p||r) = -\frac{1}{1-q} \int \mathrm{d}x\, p(x) \left[\left(\frac{r(x)}{p(x)}\right)^{1-q} - 1\right]. \tag{8.42}$$

This expression is the q-deformed relative entropy used in Tsallis' thermostatistics. It has been introduced by several authors [2, 16, 13].

The basic property of the f-divergence, with $f(u)$ a convex function satisfying $f(1) = 0$, is its positivity. Indeed, using convexity one shows

$$\begin{aligned} I(p||r) &\ge f\left(\sum_k r(k) \frac{p(k)}{r(k)}\right) \\ &= f\left(\sum_k p(k)\right) = f(1) = 0. \end{aligned} \tag{8.43}$$

The positivity of the relative entropy has been used in the Section 3.6 to discuss the *stability* properties of probability distributions belonging to the exponential family. Let us try to generalise this to q-exponential families.

One calculates

$$
\begin{aligned}
I(p||p_\theta) &= -\frac{1}{1-q}\int \mathrm{d}x\,p(x)\left[\left(\frac{p_\theta(x)}{p(x)}\right)^{1-q}-1\right]\\
&= \frac{1}{1-q} - \frac{1}{1-q}\int \mathrm{d}x\,p(x)\left(\frac{c(x)}{p(x)}\right)^{1-q}\\
&\quad\times\left[1-(1-q)\theta^k\frac{H_k(x)-\langle\langle H_k\rangle\rangle_\theta}{z(\theta)}\right]\\
&= -S_q^{\mathrm{Tsallis}}(p) + \frac{\theta^k}{z(\theta)}\int \mathrm{d}x\,p(x)\left(\frac{c(x)}{p(x)}\right)^{1-q}\left[H_k(x)-\langle\langle H_k\rangle\rangle_\theta\right]\\
&= -S_q^{\mathrm{Tsallis}}(p) + S^{\mathrm{Tsallis}}(p_\theta)\\
&\quad+\frac{\theta^k}{z(\theta)}\left[\frac{z(p)}{z(\theta)}\langle\langle H_k\rangle\rangle_p - \langle\langle H_k\rangle\rangle_\theta\right]. \qquad (8.44)
\end{aligned}
$$

Introduce

$$
\Phi_\theta^{\mathrm{ne}}(p) = S_q^{\mathrm{Tsallis}}(p) - \frac{z(p)\theta^k}{z(\theta)^2}\langle\langle H_k\rangle\rangle_p. \qquad (8.45)
$$

Then one obtains

$$
I(p||p_\theta) = \Phi_\theta^{\mathrm{ne}}(p_\theta) - \Phi_\theta^{\mathrm{ne}}(p). \qquad (8.46)
$$

In this way, one shows that the Tsallis distribution p_θ maximises the non-equilibrium Massieu functional $\Phi_\theta^{\mathrm{ne}}(p)$.

The expression (8.45) is not fully satisfactory. The thermodynamic definition of the Massieu function is $\Phi = S - \beta U$. Comparing this expression with (8.45) then suggests that the thermodynamic parameters are not the θ^k but namely $\theta^k/z(\theta)^2$, and that the thermodynamic forces are $z(p)\langle\langle H_k\rangle\rangle_p$ instead of $\langle\langle H_k\rangle\rangle_p$. In order to clarify the situation an alternative definition of relative entropy is investigated in the next Section.

8.7 Relative Entropy of the Bregman Type

The f-divergence, discussed in the previous section, is only one possible generalisation of the notion of relative entropy. An alternative definition is known as the Bregman type divergence. Given a convex function $f(u)$ with derivative $f'(u)$, it is defined by

$$
D(p||r) = \sum_j \left[f(p(j)) - f(r(j)) - (p(j)-r(j))f'(r(j))\right]. \qquad (8.47)
$$

The relation $D(p||p) = 0$ is clearly satisfied. The positivity $D(p||r) \geq 0$ is obvious if one interprets the expression in a geometric manner. The term $f(r(j)) + (p(j) - r(j))f'(r(j))$ describes a tangent in the point $r(j)$. Because

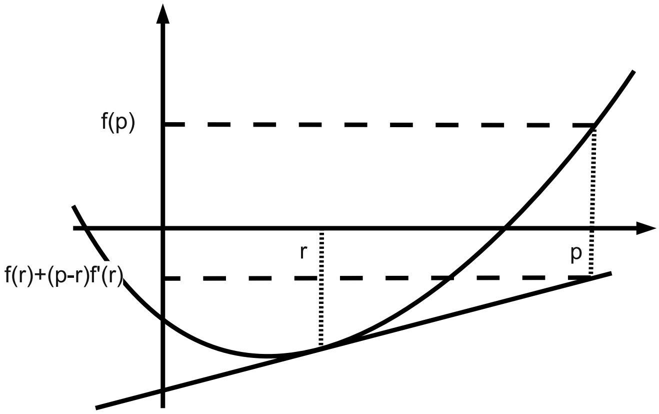

Fig. 8.2 Geometric interpretation of the Bregman divergence

$f(u)$ is convex, the tangent stays below the function value $f(u)$. See the Figure 8.2. Hence the inequality follows.

Let

$$f(u) = u \ln_q(u). \tag{8.48}$$

It is a convex function for $q < 2$. The corresponding expression for the divergence is

$$\begin{aligned} D(p||r) &= \sum_j \left[p(j)\ln_q p(j) - r(j)\ln_q r(j)\right] - s\sum_j (p(j) - r(j))\ln_q(r(j)) \\ &= -S_{2-q}^{\text{Tsallis}}(p) + S_{2-q}^{\text{Tsallis}}(r) - \frac{2-q}{1-q}\sum_j [p(j) - r(j)]r(j)^{1-q}. \end{aligned} \tag{8.49}$$

In the $q = 1$-limit this expression coincides with the relative entropy, as given by (8.39). In general the two types of divergence $D(p||r)$ and $I(p||r)$ do not coincide. In the continuous case, with weight function $c(x)$, (8.49) becomes

$$\begin{aligned} D(p||r) = &-S_{2-q}^{\text{Tsallis}}(p) + S_{2-q}^{\text{Tsallis}}(r) \\ &-\frac{2-q}{1-q}\int \mathrm{d}x\,[p(x) - r(x)]\left(\frac{r(x)}{c(x)}\right)^{1-q}. \end{aligned} \tag{8.50}$$

Introduce the alternative definition of non-equilibrium Massieu functional

$$\Phi_\theta^{\text{ne}}(p) = S_{2-q}^{\text{Tsallis}}(p) - \frac{2-q}{z(\theta)^2}\theta^k \langle H_k\rangle_p. \tag{8.51}$$

It satisfies

$$D(p||p_\theta) = \Phi_\theta^{\text{ne}}(p_\theta) - \Phi_\theta^{\text{ne}}(p). \tag{8.52}$$

Let p_θ be Tsallis distribution as given by (8.22). One calculates using (8.50)

$$\begin{aligned}
&D(p||p_\theta) + S^{\text{Tsallis}}_{2-q}(p) - S^{\text{Tsallis}}_{2-q}(p_\theta) \\
&= -\frac{2-q}{1-q}\int \mathrm{d}x\,(p(x)-p_\theta(x))\left(\frac{p_\theta(x)}{c(x)}\right)^{1-q} \\
&= -\frac{2-q}{1-q}\int \mathrm{d}x\,(p(x)-p_\theta(x))\left[1-(1-q)\alpha_0(\theta)-(1-q)\zeta^k(\theta)H_k(x)\right] \\
&= (2-q)\zeta^k(\theta)\left[\langle H_k\rangle_p - \langle H_k\rangle_\theta\right]. \qquad (8.53)
\end{aligned}$$

Hence, the definition of the non-equilibrium Massieu functional, alternative to (8.45), is

$$\Phi^{\text{ne}}(p) = S^{\text{Tsallis}}_{2-q}(p) - (2-q)\zeta^k(\theta)\langle H_k\rangle_p. \qquad (8.54)$$

Using the expression (8.23) for $\zeta^k(\theta)$ this becomes (8.51).

Box 8.2 Derivation of the relation (8.52)

See the Box 8.2. The positivity of $D(p||p_\theta)$ (assuming $0 < q < 2$) implies that $\Phi^{\text{ne}}_\theta(p)$ is maximal when $p = p_\theta$. The only differences between (8.51) and (8.45) are that the Tsallis entropy functional has index $2-q$ instead of q and that $z(p)\langle\langle H_k\rangle\rangle_p$ is replaced by $(2-q)\langle H_k\rangle_p$.

The expression (8.51) leads to an alternative interpretation of Tsallis' thermostatistics, one which is closer to the conventional wisdom of statistical physics. In this interpretation, the *thermodynamic forces* are (derivatives of) the averages $\langle H_k\rangle_p$, *not* those involving escort probabilities. The corresponding thermodynamic variables are proportional to

$$\zeta^k(\theta) = \frac{\theta^k}{z(\theta)^2} \qquad (8.55)$$

and have been introduced before - see (8.22) and the Box 8.1. The thermodynamic entropy $S(U)$ is identified with the entropy $S^{\text{Tsallis}}_{2-q}(p_\theta)$. With these changes a fully consistent formalism is obtained in which the Tsallis distribution p_θ is *thermodynamically stable* in the sense that it optimises the non-equilibrium Massieu functional as defined by (8.51). The modified expressions of the thermodynamic quantities are

$$\begin{aligned}
S(U) &= S^{\text{Tsallis}}_{2-q}(p_\theta) \\
U_k &= \langle H_k\rangle_\theta \\
\Phi(\theta) &= \Phi^{\text{ne}}_\theta(p_\theta). \qquad (8.56)
\end{aligned}$$

The formalism obtained in this way is that of the previous Chapter.

8.8 Quantum Expressions

Here, one finds quantum mechanical analogs of the main expressions of the present Section. A more profound treatment is given in the more general context of the final Chapters.

Given a density operator ρ, the Tsallis entropy is defined by

$$S_q^{\rm Tsallis}(\rho) = \frac{1}{1-q}\left(\,{\rm Tr}\,\rho^q - 1\right). \tag{8.57}$$

In the $q = 1$-limit this expression reduces to the von Neumann entropy (3.42). Note that

$$S_{2-q}^{\rm Tsallis}(\rho) = -\,{\rm Tr}\,\rho\ln_q\rho. \tag{8.58}$$

The quantum Tsallis distribution is the density operator ρ_θ given by

$$\rho_\theta = \frac{1}{Z_q(\theta)}\exp_q\left[-\theta^k\frac{H_k - \langle\langle H_k\rangle\rangle_\theta}{z(\theta)}\right], \tag{8.59}$$

with

$$z(\theta) = \,{\rm Tr}\,\rho_\theta^q \quad \text{and} \quad Z_q(\theta) = \,{\rm Tr}\,\exp_q\left[-\theta^k\frac{H_k - \langle\langle H_k\rangle\rangle_\theta}{z(\theta)}\right]. \tag{8.60}$$

The relative entropy of the Csiszár type is defined by

$$I(\rho||\sigma) = \,{\rm Tr}\,\rho(\ln_q\rho - \ln_q\sigma). \tag{8.61}$$

The relative entropy of the Bregman type is given by

$$D(\rho||\sigma) = \,{\rm Tr}\,\rho\ln_q\rho - \,{\rm Tr}\,\sigma\ln_q\sigma - \frac{2-q}{1-q}\,{\rm Tr}\,(\rho-\sigma)\sigma^{1-q}. \tag{8.62}$$

Clearly, one has $I(\rho||\rho) = D(\rho||\rho) = 0$. The proofs that $I(\rho||\sigma) \geq 0$ and $D(\rho||\sigma) \geq 0$ are not so easy because it can happen that ρ and σ do not commute. The proof of $D(\rho||\sigma) \geq 0$ will be given later on in a more general context — see the Section 11.7.

8.9 More General Entropies

A general class of entropy functionals will be discussed in Chapter 11. The entropies discussed below are somehow linked with the Tsallis entropy functional, and for that reason are mentioned already here. The Sharma-Mittal entropy functional is discussed in the next Chapter.

A two-parameter entropy was introduced in the field of non-extensive thermostatistics by Borges and Roditi [3]

$$S_{q,r}(p) = \frac{1}{q-r}\left(\sum_j p(j)^r - \sum_j p(j)^q\right). \tag{8.63}$$

One has clearly

$$S_{q,r}(p) = \frac{1}{q-r}\left[(1-r)S_r^{\text{Tsallis}} - (1-q)S_q^{\text{Tsallis}}\right]. \tag{8.64}$$

Note that $S_{q,r}(p) = S_{r,q}(p)$ and $S_{q,1}(p) = S_q^{\text{Tsallis}}(p)$.

The entropy functional (8.63) generalizes the entropy $S_{q,q^{-1}}(p)$, introduced by Abe [1]

$$S_{q^{-1},q}(p) = \frac{1}{q-q^{-1}}\left(\sum_j p(j)^{q^{-1}} - \sum_j p(j)^q\right). \tag{8.65}$$

It also generalizes the entropy $S_\kappa(p)$ introduced by Kaniadakis [7]

$$S_{q,r}(p) = \frac{1}{2\kappa}\left(\sum_j p(j)^{1-\kappa} - \sum_j p(j)^{1+\kappa}\right). \tag{8.66}$$

Landsberg and Vedral [8] introduced the following variation on Tsallis' entropy

$$S(p) = \frac{1}{1-q}\,\frac{\sum_j p(j)^q - 1}{\sum_j p(j)^q} = \frac{1}{1-q}\left(1 - \frac{1}{\sum_j p(j)^q}\right). \tag{8.67}$$

This entropy has been further studied in [12, 18, 19].

Problems

8.1. Conventional energy constraints

Assume that $0 < q < 2$. Show that the Tsallis entropy $S_q^{\text{Tsallis}}(p)$ is maximised by a probability distribution of the form

$$p_\theta(j) = \exp_q^*(-\alpha(\theta) - \theta^k H_k(j)) \tag{8.68}$$

when the constraints are of the conventional type $\langle H_k \rangle_p = U_k$, instead of those involving the escort probability. Remember that the dual exponential function $\exp_q^*$ is defined by (7.10).

8.2. Identity
Let be given a probability distribution $p(j)$, together with its escort $P(j)$, defined by (7.24). Prove that the Tsallis entropy functional satisfies the identity (see Theorem 13 of [14])

$$\exp_q \left(S_q^{\mathrm{Tsallis}}(p) \right) = \exp_{1/q} \left(S_{1/q}^{\mathrm{Tsallis}}(P) \right). \tag{8.69}$$

Notes

There is a huge literature about Tsallis' thermostatistics, most of which is not covered here.

There have been different versions of Tsallis' thermostatistics. The Tsallis probability distribution, presented here (8.17), is that of the third version of the theory, proposed in [17]. This paper claims that the parameter θ, appearing in (8.17) when only one Hamiltonian $H(j)$ is involved, coincides with the inverse temperature β and justifies this claim by citing [11]. In my opinion, this claim is not justified because the Tsallis distribution falls out of the scope of [11].

The mean field character of non-extensive thermostatistics was noted in [9]. It explains why Tsallis' thermostatistics exhibits thermodynamic instabilties in situations where they are totally unexpected. These can be understood as being consequences of the Tsallis distribution belonging to a curved q-exponential family. The parameter θ is an *effective temperature*. After reparametrisation

The Tsallis distribution has been considered as a parametrised family in the recent work of Campisi and Bagci [4]. They study the formalism in the context of Boltzmann's concept of orthodes.

The definition (8.49) of the q-deformed relative entropy, is a special case of that introduced in [10] and discussed in the Chapter 11. See also the Problem 11.1 of the latter Chapter.

Objectives

- Know the Tsallis entropy functional and its properties.
- Know the Tsallis distribution.
- Show that the Tsallis distribution with deformation parameter q belongs to the curved q-exponential family.

- Know about the existence of two different notions of q-deformed relative entropy, one of the Csiszár type, the other of the Bregman type.
- Be aware that unexpected thermodynamic instabilities may occur in Tsallis' thermostatistics as found in the literature. Know how to cure these problems.
- Replace the classical expressions by their quantum analogs.

References

1. Abe, S.: A note on the q-deformation-theoretic aspect of the generalized entropies in nonextensive physics. Phys. Lett. A **224**, 326–330 (1997) 128
2. Abe, S.: q-deformed entropies and Fisher metrics. In: P. Kasperkovitz, D. Grau (eds.) Proceedings of The 5th International Wigner Symposium, August 25–29, 1997, Vienna, Austria, p. 66. World Scientific, Singapore (1998) 123, 177
3. Borges, E., Roditi, I.: A family of nonextensive entropies. Phys. Lett. A **246**, 399–402 (1998) 128, 176
4. Campisi, M., Bagci, G.B.: Tsallis ensemble as an exact orthode. Phys. Lett. A **362**, 11–15 (2007) 129
5. Daróczy, Z.: Generalized information functions. Inform. Control **16**, 36–51 (1970) 115
6. Havrda, J., Charvat, F.: Quantification method of classification processes: concept of structural α-entropy. Kybernetica **3**, 30–35 (1967) 115
7. Kaniadakis, G., Scarfone, A.: A new one parameter deformation of the exponential function. Physica A **305**, 69–75 (2002) 128, 152, 176
8. Landsberg, P., Vedral, V.: Distributions and channel capacities in generalized statistical mechanics. Phys. Lett. A **247**, 211–217 (1998) 128
9. Naudts, J.: Generalized thermostatistics and mean field theory. Physica A **332**, 279–300 (2004) 129, 162
10. Naudts, J.: Continuity of a class of entropies and relative entropies. Rev. Math. Phys. **16**, 809822 (2004); Errata. Rev. Math. Phys. **21**, 947–948 (2009) 129, 162, 177
11. Plastino, A., Plastino, A.: On the universality of thermodynamics' Legendre transform structure. Phys. Lett. A **226**, 257–263 (1997) 129
12. Rajagopal, A., Abe, S.: Implications of form invariance to the structure of nonextensive entropies. Phys. Rev. Lett. **83**, 1711–1714 (1999) 128
13. Shiino, M.: H-theorem with generalized relative entropies and the Tsallis statistics. J. Phys. Soc. Jpn. **67(11)**, 3658–3660 (1998) 123
14. Suyari, H., Wada, T.: Multiplicative duality, q-triplet and (μ, ν, q)-relation derived from the one-to-one correspondence between the (μ, ν)-multinomial coefficient and Tsallis entropy S_q. Physica A **387**, 71–83 (2008) 129
15. Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. **52**, 479–487 (1988) 115
16. Tsallis, C.: Generalized entropy-based criterion for consistent testing. Phys. Rev. E **58**, 1442–1445 (1998) 123
17. Tsallis, C., Mendes, R., Plastino, A.: The role of constraints within generalized nonextensive statistics. Physica A **261**, 543–554 (1998) 112, 117, 129
18. Yamano, T.: Information theory based on nonadditive information content. Phys. Rev. E **63**, 046105 (2001) 128
19. Yamano, T.: Source coding theorem based on a nonadditive information content. Physica A **305**, 190–195 (2002) 128

Chapter 9
Changes of Scale

9.1 Kolmogorov-Nagumo Averages

Experimental data are often plotted on logarithmic paper. In particular, if the data obey a power law relation

$$y_i \simeq \frac{A}{x_i^\alpha}, \tag{9.1}$$

then

$$\ln y_i \simeq \ln A - \alpha \ln x_i. \tag{9.2}$$

Hence, on the logarithmic paper the data points fall roughly on a straight line. By linear regression one can then estimate the constants $\ln A$ and α. Note however that linear regression on (9.2) does not produce the same results as fitting (9.1) directly to the data using non-linear fitting software. The reason is that the experimental errors are weighed differently. This means that the procedure of estimating parameters is sensitive to the scale on which the experimental data are presented. It is therefore of interest to study the role of scale changes in a systematic manner.

The only requirements for a *scaling function* $\phi(u)$ are that it is continuous and strictly monotonic. It can be either increasing or decreasing. But only increasing functions are considered here because this is the obvious choice. Given a scaling function $\phi(u)$, and a discrete probability distribution $p_1, p_2, \cdots p_n$, the *Kolmogorov-Nagumo average* of the sequence of numbers $x_1, x_2, \cdots x_n$ is defined by

$$\langle x \rangle_\phi = \phi^{-1}\left(\sum_{k=1}^n p_k \phi(x_k)\right). \tag{9.3}$$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_9,

The data points x_k are scaled with the function $\phi(u)$. Next, the average is calculated using the probabilities p_k. Finally, the average is scaled back using the inverse function $\phi^{-1}(u)$.

Two basic properties of linear averages still hold for Kolmogorov-Nagumo averages.

· If all x_k have the same value x then also the average $\langle x\rangle_\phi$ has the value x.
· If $x_k \leq y_k$ for all k then $\langle x\rangle_\phi \leq \langle y\rangle_\phi$. In particular, if $0 \leq y_k$ then $0 \leq \langle y\rangle_\phi$.

9.2 Rényi's Alpha Entropies

The Kolmogorov-Nagumo averages have been used by Alfréd Rényi [2] to introduce a family of entropy functionals, called the *alpha entropies*, or *Rényi entropies*. The index q is used instead of α to stress the link with the Tsallis entropy functional S_q^{Tsallis}, but also to avoid a conflict of notation in the Section on fractals. The definition is

$$S^q(p) = \frac{1}{1-q} \ln \left(\sum_j p(j)^q \right). \tag{9.4}$$

It is a decreasing function of q. See the Box 9.1.

Here we show that the Rényi entropy functional $S^q(p)$ is a decreasing function of the parameter q.

One calculates

$$(1-q)^2 \left(\sum p(i)^q\right) \frac{\mathrm{d}S^q}{\mathrm{d}q} = f\left(\sum p(i)^q\right) + (1-q)\sum p(i)^q \ln p(i), \tag{9.5}$$

with $f(x) = x\ln x$. The function $f(x)$ is convex. Hence one has

$$\begin{aligned} f\left(\sum p(i)^q\right) &\leq \sum p(i) f(p(i)^{q-1}) \\ &= \sum p(i)^q \ln p(i)^{q-1} \\ &= (q-1)\sum p(i)^q \ln p(i). \end{aligned} \tag{9.6}$$

This shows that (9.5) cannot be positive.

Box 9.1 q-dependence of Rényi's entropy functional

The Rényi entropy functional can be written as a Kolmogorov-Nagumo average

$$S^q(p) = \left\langle \ln \frac{1}{p} \right\rangle_{\phi_q} \tag{9.7}$$

with $\phi_q(x) = \ln_q e^x$. Indeed, one evaluates

$$\begin{aligned}
\left\langle \ln \frac{1}{p} \right\rangle_{\phi_q} &= \phi_q^{-1}\left(\sum_j p(j)\phi_q[\ln \frac{1}{p(j)}]\right) \\
&= \phi_q^{-1}\left(\sum_j p(j) \ln_q \frac{1}{p(j)}\right) \\
&= \phi_q^{-1}\left(\frac{1}{1-q}\sum_j p(j)\left[p(j)^{q-1} - 1\right]\right) \\
&= \ln \exp_q\left(\frac{1}{1-q}\left[\sum_j p(j)^q - 1\right]\right) \\
&= \ln \left(\sum_j p(j)^q\right)^{1/(1-q)} \\
&= \frac{1}{1-q} \ln \left(\sum_j p(j)^q\right) \\
&= S^q(p).
\end{aligned} \tag{9.8}$$

It is also clear that in the limit $q = 1$ this entropy reduces to the Boltzmann-Gibbs-Shannon entropy. Indeed, the scaling function becomes linear $\phi_1(x) = x$ so that

$$\begin{aligned}
S^1(p) &= \left\langle \ln \frac{1}{p} \right\rangle \\
&= -\sum_j p(j) \ln p(j) \\
&= S^{\mathrm{BGS}}(p).
\end{aligned} \tag{9.9}$$

9.3 Rényi or Tsallis?

There is a straightforward relation between Rényi's $S^q(p)$ and the *Tsallis' entropy functional* $S_q^{\mathrm{Tsallis}}(p)$. Indeed, one has

Here we show that Rényi's entropy functional $S^q(p)$ is additive.
Let $p(i,j) = p_A(i)p_B(j)$. Then one has

$$\begin{aligned} S^q(p) &= \frac{1}{1-q} \ln \left(\sum_{i,j} p(i,j)^q \right) \\ &= \frac{1}{1-q} \ln \left(\sum_i p_A(i)^q \right) \left(\sum_j p_B(j)^q \right) \\ &= \frac{1}{1-q} \ln \left(\sum_i p_A(i)^q \right) + \frac{1}{1-q} \ln \left(\sum_j p_B(j)^q \right) \\ &= S^q(p_A) + S^q(p_B). \end{aligned} \tag{9.10}$$

This shows the additivity.

Box 9.2 Additivity of Rényi's entropy

$$S^q(p) = \frac{1}{1-q} \ln \left(1 + (1-q) S_q^{\text{Tsallis}}(p)\right) = \xi \left[S_q^{\text{Tsallis}}(p)\right], \tag{9.11}$$

with

$$\xi(u) = \frac{1}{1-q} \ln(1 + (1-q)u) \tag{9.12}$$

Note that the function $\xi(u)$ is well-defined when $1 + (1-q)u > 0$ and is strictly increasing because

$$\xi'(u) = \frac{1}{1 + (1-q)u} > 0. \tag{9.13}$$

Hence, when the entropy is maximized in a variational principle, it cannot make any difference whether the Tsallis entropy functional is used or Rényi's, because when one of the two reaches its maximum then also the other is maximal. However, the thermodynamic entropy $S(U)$ is a concept which differs from that of the entropy functional. This raises the question whether $S(U)$ equals $S^q(p)$ or $S_q^{\text{Tsallis}}(p)$, when evaluated for the probability distribution p which optimizes the entropy functional under the constraint that the average energy equals U. This question has already been touched upon in the previous Chapter. Here two arguments are given which indicate that Rényi's entropy functional is better suited for this purpose.

A well-known property of Rényi's entropy functional is that it is *additive*, while Tsallis' entropy functional is not. This means that when the system is composed of two independent subsystems A and B, and consequently, the probability distribution p is the product of the distributions p_A and p_B, then the entropy $S^q(p)$ is the sum of $S^q(p_A)$ and $S^q(p_B)$. See the Box 9.2. Now, additivity of the thermodynamic entropy $S(U)$ is one of the corner stones of thermodynamics. hence it is clear that Rényi's entropy functional is better suited for this role.

A further argument in favour of Rényi's entropy functional comes from the properties that one expects the temperature to hold. Let us assume that the *temperature* T is defined by the thermodynamic relation

$$\frac{1}{T} = \frac{\mathrm{d}S}{\mathrm{d}U}. \tag{9.14}$$

It is then clear that a different result for T will be obtained whether Tsallis' or Rényi's entropy functional is used. Both options are now compared in an example.

9.4 Configurational Temperature

The monoatomic gas has been treated already in the Section 7.8. Here we show that the temperature of the configurational subsystem equals that of the kinetic energy subsystem provided that Rényi's entropy is used.

It was shown already in the Section 7.8 that the configurational probability distribution $q_E^{\mathrm{conf}}(\mathbf{q})$ belongs to the q-exponential family, where $q = 1 - \dfrac{2}{3N-2}$ is determined by the number of particles N. More precisely, it can be written as

$$q_E^{\mathrm{conf}}(\mathbf{q}) = c_N \exp_q(-\alpha(E) - \theta\mathcal{V}(\mathbf{q})) \tag{9.15}$$

with

$$\alpha(E) = \frac{1}{1-q}\left[1 - \left(\frac{2ma^2}{h^2}\right)^{2-q}\frac{E}{\omega(E)^{1-q}}\right] \tag{9.16}$$

and with $c_N = \dfrac{3N}{2}B_{3N}(1)$, where $B_n(1)$ is the volume of a unit sphere in n dimensions. The function $\mathcal{V}(\mathbf{q})$ is the potential energy term of the Hamiltonian. The function $\omega(E)$ is the density of states as a function of the total energy E. The parameter θ depends on E via

$$\theta = \frac{1}{1-q}\left(\frac{2ma^2}{h^2}\right)^{2-q}\omega(E)^{q-1}. \tag{9.17}$$

The constants a and h are arbitrary and are introduced for dimensional reasons.

The pdf (9.15) maximises the entropy functional $I(p)$, given by (7.34), and which coincides with $S_{2-q}^{\mathrm{Tsallis}}(p)$. Let us therefore calculate the Rényi entropy

$$\begin{aligned} S^{2-q}(q_E^{\text{conf}}) &= \frac{1}{q-1} \ln \frac{1}{N!a^{3N}} \int \mathrm{d}\mathbf{q}\, q_E^{\text{conf}}(\mathbf{q}) \left[\frac{q_E^{\text{conf}}(\mathbf{q})}{c_N} \right]^{1-q} \\ &= \frac{1}{q-1} \ln \frac{1}{N!a^{3N}} \int \mathrm{d}\mathbf{q}\, q_E^{\text{conf}}(\mathbf{q}) \left[1-(1-q)\alpha(E) - (1-q)\theta\mathcal{V}(\mathbf{q}) \right] \\ &= \frac{1}{q-1} \ln \left[1-(1-q)\alpha(E) - (1-q)\theta\langle\mathcal{V}\rangle_\theta \right] \\ &= -\left(\frac{3N}{2} - 1 \right) \ln \frac{U^{\text{kin}}}{\epsilon^{2-q}\omega(E)^{1-q}}, \end{aligned} \tag{9.18}$$

where $U^{\text{kin}} = E - U^{\text{conf}}$ is the average kinetic energy, $U^{\text{conf}} = \langle\mathcal{V}\rangle_\theta$ is the average potential energy, and $\epsilon = h^2/2ma^2$ is a constant with the dimensions of an energy.

Assume now that the Rényi entropy (9.18) coincides with the thermodynamic entropy $S^{\text{conf}}(U^{\text{conf}})$ of the configurational subsystem. Then the configurational temperature is calculated as follows

$$\begin{aligned} \frac{1}{T} &= \frac{\mathrm{d}S^{\text{conf}}}{\mathrm{d}U^{\text{conf}}} \\ &= \left(\frac{3N}{2} - 1 \right) \frac{1}{U^{\text{kin}}} \left[1 - \frac{\mathrm{d}E}{\mathrm{d}U^{\text{conf}}} \right] - \frac{\omega'(E)}{\omega(E)} \frac{\mathrm{d}E}{\mathrm{d}U^{\text{conf}}}. \end{aligned} \tag{9.19}$$

The derivative $\mathrm{d}E/\mathrm{d}U^{\text{conf}}$ can be obtained from the relation

$$E - U^{\text{conf}} = U^{\text{kin}} = \frac{3N}{2} \frac{\Omega(E)}{\omega(E)} \tag{9.20}$$

— see (4.14) and (7.70). This relation implies

$$\frac{\mathrm{d}U}{\mathrm{d}E} = 1 - \frac{3N}{2} + U^{\text{kin}} \frac{\omega'(E)}{\omega(E)}. \tag{9.21}$$

Expression (9.19) now simplifies to

$$\frac{1}{T} = \frac{\omega(E)}{\Omega(T)}. \tag{9.22}$$

But this is precisely the expression (4.14) for the temperature T as obtained from the modified Boltzmann entropy (4.12)— note that here $k_{\mathrm{B}} = 1$ has been used. It also coincides with the temperature of the subsystem of kinetic energies (assigning $\frac{1}{2}k_{\mathrm{B}}T$ to each degree of freedom). One therefore concludes that the use of Rényi's entropy functional for the configurational subsystem leads to a value of the *configurational temperature* which coincides with the one of the total system and the one derived from the speed of the atoms.

9.5 Fractal Dimensions

Let us now consider an important application of Rényi's entropy functional, with the intention to find out whether there is any relation with the main topic of this book, which is the (deformed) exponential family. It will turn out that this is not the case, at least not in an obvious manner..

The Rényi entropies have been introduced in physics in the context of multifractals. A *fractal* subset of $\mathbf{R}^d$ has a vanishing (hyper)volume in d dimensions, while the volume becomes finite in some non-integer dimension ν less than d. This finite volume is then related to the linear size l of the set as l^α. This exponent α says how the volume *scales* with the linear size of the set.

A related concept is that of fractal measures, see the Box 9.3 for an example. For simplicity of notations only fractal measures on the interval $[0,1]$ are considered. Also, the *coarse graining* of the interval $[0,1]$ is limited to the regular partition into 2^n intervals of equal length. More general partitions could be considered and can simplify the applications. This is not done here because this would obscure the exposition.

The fractal measure p on the interval $[0,1]$ determines a discrete probability distribution $p^{(n)}$ by

$$p^{(n)}(j) = \int_{j2^{-n}}^{(j+1)2^{-n}} \mathrm{d}p(x). \tag{9.23}$$

Assume that the Rényi entropies $S^q(p^{(n)})$ are extensive quantities in the sense that they diverge linearly with n. Then the *Rényi dimension* (also called the *generalised dimension*) D_q of p is defined by

$$D_q(p) = \lim_{n\to\infty} \frac{1}{n\ln 2} S^q(p^{(n)}). \tag{9.24}$$

The Rényi entropy of a discrete probability distribution cannot be negative. Hence, the dimension satisfies $D_q(p) \geq 0$. It is a decreasing function of q. See the Figure 9.1 for an example. The value at $q=0$ coincides with what is known in the literature as the *box dimension* of the fractal set which supports the probability distribution. It usually coincides with the *Hausdorff dimension,* although the latter is defined in a different manner, involving more partitions of $[0,1]$ than only those in intervals of equal length.

The limit of large n in (9.24) looks like a *thermodynamic limit.* This resemblance is made hard in what is known as *multifractal analysis.* Its main result is that (9.24) can be rewritten as

$$\tau(q) \equiv (q-1)D_q(p) = \inf_\alpha \left(q\alpha - f(\alpha)\right), \tag{9.29}$$

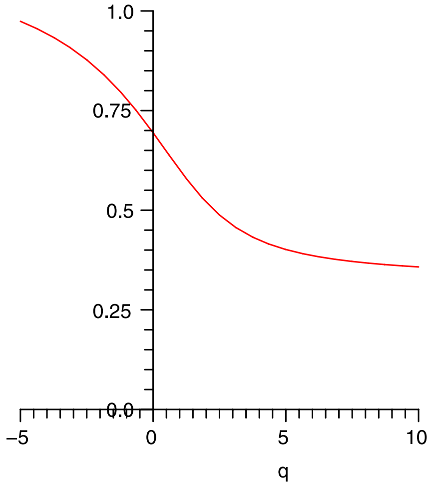

Fig. 9.1 The Rényi dimension D_q of the asymmetric Cantor set of the Box 9.3, with $l_1 = 1/2$. $l_2 = 1/4$, and $a = 0.8$

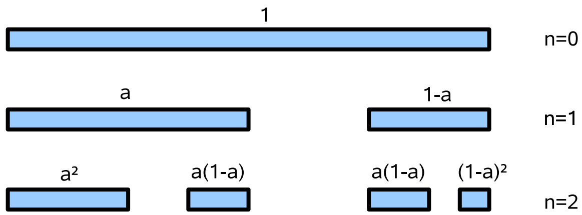

Fig. 9.2 Construction of a Cantor set with $l_1 = 0.5$ and $l_2 = 0.25$. Probabilities are assigned with $p_1 = a$ and $p_2 = 1 - a$

for some function $f(\alpha)$. This expression enforces the analogy with the thermodynamic formalism — see (3.15). The parameter α plays the role of the energy, the parameter q that of the inverse temperature β. Hence, $\tau(q)$ is minus the *Massieu function*, and is proportional to the *free energy*. The function $f(\alpha)$ then corresponds with the *thermodynamic entropy* $S(U)$. Further support for this identification follows when the *escort probability distribution* is introduced. Let

$$p_q^{(n)}(j) = \frac{1}{z^{(n)}(q)} \left(p^{(n)}(j)\right)^q , \tag{9.30}$$

with

The fractal measure p $(0 < a < 1)$ with support in the asymmetric *Cantor set* is constructed as a limit of probability distributions $r_n(x)$. It starts with the interval $[0,1]$ at level $n = 0$. The corresponding probability distribution is $r_0(x) = 1$, uniformly. Next divide the interval in two pieces $[0, l_1]$ and $[1-l_2, 1]$, where l_1 and l_2 are positive constants satisfying $l_1 + l_2 < 1$. The corresponding probability distribution is

$$\begin{aligned} r_1(x) &= \frac{a_1}{l_1} \quad \text{if } 0 < x < l_1, \\ &= 0 \quad \text{if } l_1 < x < 1 - l_2 \\ &= \frac{a_2}{l_2} \quad \text{if } 1 - l_2 < x < 1. \end{aligned} \tag{9.25}$$

The probabilities a_1 and a_2 satisfy $a_1 = a$ and $a_2 = 1 - a$. This division step is repeated infinitely many times — See the Figure 9.2. At level n the probability distribution $r_n(x)$ gives probability $a_1^m a_2^{n-m}$ to $\binom{n}{m}$ intervals of length $l_1^m l_2^{n-m}$. One verifies that

$$\int_0^1 \mathrm{d}x\, r_n(x) = \sum_{m=0}^{n} \binom{n}{m} a_1^m a_2^{n-m} = 1. \tag{9.26}$$

The fractal measure $p(x)$ is now defined as the limit of $r_n(x)$ for n tending to infinity.

Next consider a partition of $[0,1]$ into 2^n intervals of equal length. Assume $l_1 = 1/2$ and $l_2 = 1/4$ for simplicity. Then the probability of the interval $[j2^{-n}, (j+1)2^{-n}]$ is fixed during the above construction at a level n' depending on j and does not anymore change when n' increases. This makes it possible to write the following recursion relation

$$z^{(n+1)}(q) = a^q z^{(n)}(q) + (1-a)^q z^{(n-1)}(q) \tag{9.27}$$

with $z^{(n)}(q) = \exp\left[(1-q)S^q(p^{(n)})\right]$. By induction one then shows that

$$S^q(p^{(n)}) = \frac{1}{1-q} \ln \sum_{m=0}^{n} \sum_{l=0}^{2l+m\leq n+1} \binom{n-l}{m} a^{qm}(1-a)^{ql}. \tag{9.28}$$

For $q = 0$, the recursion relation (9.27) defines the *Fibonacci numbers* $1, 2, 3, 5, 8, 13, \cdots$. From their asymptotics one derives $D_0(p) = \ln((1+\sqrt{5})/2)/\ln 2$, which is approximately 0.694.

Box 9.3 Example of a fractal measure

$$z^{(n)}(q) = \sum_j \left(p^{(n)}(j)\right)^q. \tag{9.31}$$

Then the Rényi entropy becomes

$$S^q(p^{(n)}) = \frac{1}{1-q} \ln z^{(n)}(q). \tag{9.32}$$

Note that the escort probability is used here to turn the probability distribution $p^{(n)}(j)$ into a one-parameter family of probability distributions, with q being this parameter. This is conceptually different from its use in the previous Chapters, where q is fixed and refers to the q-exponential family to which the probability distribution belongs.

By construction, $p_q^{(n)}(j)$ belongs to the *exponential family.* Indeed, one can write

$$p_q^{(n)}(j) = \exp\left(-\Phi^{(n)}(q) - qH^{(n)}(j)\right) \tag{9.33}$$

with

$$\Phi^{(n)}(q) = \ln z^{(n)}(q) \quad \text{and} \quad H^{(n)}(j) = -\ln p^{(n)}(j). \tag{9.34}$$

From the *variational principle* of Chapter 3 then follows that

$$\begin{aligned}(q-1)S^q(p^{(n)}) &= -\Phi^{(n)}(q) \\ &= \inf_p \{q\langle H^{(n)}\rangle_p - S^{\mathrm{BGS}}(p)\}.\end{aligned} \tag{9.35}$$

This expression is already similar to (9.29), but is not yet the same.

9.6 Microcanonical Description

Up to here the description of fractal measures corresponds to that of a canonical ensemble. Let us now introduce a microcanonical description. Write (9.31) as

$$z^{(n)}(q) = \int \mathrm{d}\alpha\, \rho^{(n)}(\alpha) e^{-nq\alpha \ln 2}, \tag{9.36}$$

with

$$\rho^{(n)}(\alpha) = \sum_{j=0}^{2^n-1} \delta\left(\alpha + \frac{\ln p^{(n)}(j)}{n \ln 2}\right). \tag{9.37}$$

The expression (9.36) is the analogue of (5.9). A corresponding microcanonical distribution is

$$\begin{aligned}q_\alpha^{(n)}(j) &= \frac{1}{c^{(n)}(\alpha)} \quad \text{if } \ln p^{(n)}(j) = -n\alpha \ln 2, \\ &= 0 \quad \text{otherwise},\end{aligned} \tag{9.38}$$

where $c^{(n)}(\alpha)$ is the number of j for which $\ln p^{(n)}(j) = -n\alpha \ln 2$ holds. Indeed, one has now

$$
\begin{aligned}
&\frac{1}{z^{(n)}(q)}\int \mathrm{d}\alpha\,\rho^{(n)}(\alpha)e^{-nq\alpha\ln 2}q_\alpha^{(n)}(j)\\
&=\frac{1}{z^{(n)}(q)}\sum_{l=0}^{2^n-1}\int \mathrm{d}\alpha\,\delta\left(\alpha+\frac{\ln p^{(n)}(l)}{n\ln 2}\right)e^{-nq\alpha\ln 2}q_\alpha^{(n)}(j)\\
&=\frac{1}{z^{(n)}(q)}\sum_{l=0}^{2^n-1}\left(p^{(n)}(l)\right)^q\int \mathrm{d}\alpha\,\delta\left(\alpha+\frac{\ln p^{(n)}(l)}{n\ln 2}\right)q_\alpha^{(n)}(j)\\
&=\sum_{l=0}^{2^n-1}p_q^{(n)}(l)\int \mathrm{d}\alpha\,\delta\left(\alpha+\frac{\ln p^{(n)}(l)}{n\ln 2}\right)q_\alpha^{(n)}(j)\\
&=p_q^{(n)}(j).
\end{aligned}
\tag{9.39}
$$

This is the decomposition of $p_q^{(n)}(j)$ into the microcanonical distributions $q_\alpha^{(n)}(j)$.

From (9.35) now follows

$$(q-1)S^q(p^{(n)})\le nq\alpha\ln 2-\ln c^{(n)}(\alpha). \tag{9.40}$$

Here we used that $\langle H^{(n)}\rangle_{q_\alpha^{(n)}}=n\alpha\ln 2$ and $S^{\mathrm{BGS}}(q_\alpha^{(n)})=\ln c^{(n)}(\alpha)$. Let

$$f(\alpha)=\lim_n\frac{\ln c^{(n)}(\alpha)}{n\ln 2}. \tag{9.41}$$

Then (9.40) becomes

$$\tau(q)\le q\alpha-f(\alpha). \tag{9.42}$$

This proves part of (9.29).

It remains to show that there exists an α for which the equality is reached. This is the value of α which maximises the hyperdistribution

$$\frac{1}{z^{(n)}(q)}\rho^{(n)}(\alpha)e^{-nq\alpha\ln 2}. \tag{9.43}$$

However, note that $\rho^{(n)}(\alpha)$ is a singular function. This makes the proof non-trivial. The reader is referred to [13]. A numerical verification in the context of the asymmetric Cantor set is discussed in the Box 9.4.

Use the notations of the Box 9.3, with $l_1 = 1/2$ and $l_2 = 1/4$. There are $\binom{n-l}{m}$ intervals of length 2^{-n} with probability $a^m(1-a)^l$ — see (9.28). They contribute to $c^{(n)}(\alpha)$ with α given by

$$\alpha \ln 2 = -x \ln a - y \ln(1-a), \tag{9.44}$$

with $x = m/n$ and $y = l/n$. Using Stirling's approximation,

$$\ln a! \sim \frac{1}{2} \ln 2\pi a + a \ln a - a + \frac{1}{12a} + \cdots, \tag{9.45}$$

the contribution to $f(\alpha)$ equals

$$\begin{aligned} f(\alpha) \ln 2 &= \lim_n \frac{1}{n} \ln \sum \binom{n-l}{m} \\ &= \max\left[(1-y)\ln(1-y) - x \ln x - (1-x-y)\ln(1-x-y)\right]. \end{aligned} \tag{9.46}$$

The sum in the first line and the maximum in the second line are restricted to x and y satisfying (9.44). This is at the border of the index range when $2l+m$ equals n or $n+1$. Working this out gives

$$\begin{aligned} x &= 1 - 2y \\ y &= \frac{\alpha \ln 2 + \ln a}{2 \ln a - \ln(1-a)}. \end{aligned} \tag{9.47}$$

The function $q\alpha - f(\alpha)$ is minimal when $q = f'(\alpha)$. This gives

$$q = \frac{2\ln(1-2y) - \ln y - \ln(1-y)}{2 \ln a - \ln(1-a)}. \tag{9.48}$$

Inverting this relation gives $y = \frac{1}{2}(1 - \frac{a^q}{w})$ with $w = \sqrt{a^{2q} + 4(1-a)^q}$. One finally obtains

$$\tau(q) \ln 2 = (q\alpha - f(\alpha) \ln 2 = -\ln w - \ln(1-y). \tag{9.49}$$

This quantity, divided by $(q-1)\ln 2$, has been plotted in the Figure 9.1. Its agreement with the result obtained from (9.28) has been verified numerically.

Box 9.4 Microcanonical treatment of the asymmetric Cantor set

9.7 Sharma-Mittal Entropy Functional

Sharma and Mittal [14] introduced the two-parameter family of entropy functions

$$S^{\mathrm{SM}}_{qr}(p) = \frac{1}{1-r}\left(\left[\sum_j p(j)^q\right]^{(r-1)/(q-1)} - 1\right). \tag{9.51}$$

The parameters q and r are assumed to be positive and different from 1. Tsallis' entropy functional is recovered when $r = q$. Rényi's q-entropy follows in the limit $r = 1$.

The entropy functional of Sharma and Mittal can be written as a Kolmogorov-Nagumo average. Indeed, one has [5]

$$S^{\mathrm{SM}}_{qr}(p) = \left\langle \ln_r\left(\frac{1}{p}\right)\right\rangle_\phi \tag{9.52}$$

with

$$\begin{aligned}\phi(x) &= \ln_q \exp_r(x)\\ &= \frac{1}{1-q}\left([1+(1-r)x]_+^{(1-q)/(1-r)} - 1\right).\end{aligned} \tag{9.53}$$

To see (9.52), one calculates

$$\begin{aligned}\left\langle \ln_r\left(\frac{1}{p}\right)\right\rangle_\phi &= \ln_r \exp_q\left(\sum_j p(j)\ln_q \exp_r \ln_r\left(\frac{1}{p}\right)\right)\\ &= \ln_r \exp_q\left(\frac{1}{1-q}\sum_j p(j)\left[p(j)^{q-1}-1\right]\right)\\ &= \ln_r \exp_q\left(\ln_q\left[\sum_j p(j)^q\right]^{1/(1-q)}\right)\\ &= \ln_r\left(\left[\sum_j p(j)^q\right]^{1/(1-q)}\right)\\ &= \frac{1}{1-r}\left(\left[\sum_j p(j)^q\right]^{(1-r)/(1-q)} - 1\right)\\ &= S^{\mathrm{SM}}_{qr}(p).\end{aligned} \tag{9.50}$$

Box 9.5 Derivation of (9.52)

See the Box 9.5. Note that Tsallis' entropy, which requires $r = q$, involves a linear average instead of a Kolmogorov-Nagumo average. From the calculation (9.50) in the Box 9.5 follows also that

$$S_{qr}^{\mathrm{SM}}(p) = \ln_r \exp_q S_q^{\mathrm{Tsallis}}(p). \tag{9.54}$$

Since both $\ln_r(x)$ and $\exp_q(x)$ are strictly increasing functions, it follows that S_{qr}^{SM} increases if and only if S_q^{Tsallis} increases. Hence, a probability distribution which maximises S_q^{Tsallis} also maximises S_{qr}^{SM} and vice-versa. This will be used below.

Consider now the problem of optimising the Sharma-Mittal entropy $S_{qr}^{\mathrm{SM}}(p)$ with the constraint that the non-linear average energy $\langle H\rangle_\phi$ has some given value U. Here, $\phi(x)$ still equals $\ln_q \exp_r(x)$. Because the inverse function $\phi^{-1}(x) = \ln_r \exp_q(x)$ is strictly increasing, the problem is equivalent with optimising

$$S_q^{\mathrm{Tsallis}}(p) \equiv \sum_j p(j)\, \ln_q \frac{1}{p(j)} \tag{9.55}$$

with constraint

$$\sum_j p(j)\phi(\beta_0 H(j)) = \phi(\beta_0 U). \tag{9.56}$$

The constant β_0 has been inserted for dimensional reasons. The solution to this problem is of the form (see (8.1))

$$p_\beta(j) = \exp_q^* \left(-\Phi(\beta) - \beta\phi(\beta_0 H(j))\right), \tag{9.57}$$

where $\Phi(\beta)$ is the normalisation, and where β should be adjusted so that (9.56) is satisfied.

Note that (9.57) in the limit $r = 1$ becomes

$$p_\beta(j) = \exp_q^* \left(-\Phi(\beta) + \frac{\beta}{q-1}\left[e^{-(q-1)\beta_0 H(j)} - 1\right]\right). \tag{9.58}$$

This coincides with the probability distribution postulated in [15] on an ad hoc basis in order to improve theoretical fits to experimental protein-folding data. On the other hand, in the limit $q = 0$ expression (9.57) becomes

$$p_\beta(j) = \frac{1}{\left[1 + \Phi(\beta) - \beta + \beta[1 + (1-r)\beta_0 H(j)]_+^{1/(1-r)}\right]_+}. \tag{9.59}$$

This expression is used in the example below.

9.8 Zipf's Law

The following application belongs to the domain of quantitative linguistics. Consider the complete works of Shakespeare. These are freely available in electronic version on the Internet [1], together with the works of many other authors. This makes it easy to do the following experiment. Count how many times the same word occurs in the collected body of text. Next make a ranking of the most frequently used words and plot their frequency f versus their ranking r on semi-logarithmic paper. This gives a plot like that of Figure 9.3, based on data from a corpus of 2606 books in English, containing 448,359 different words (see [12]).

Zipf, in 1932, observed that curves $f(r)$ obtained in this way decay by a powerlaw

$$f(r) \simeq \frac{A}{r^\alpha} \tag{9.60}$$

with exponent α slightly larger than 1. Benoit Mandelbrot used this in 1966 as an example of fractal behaviour in society. He proposed [9] the more general relation, involving an additional fitting parameter,

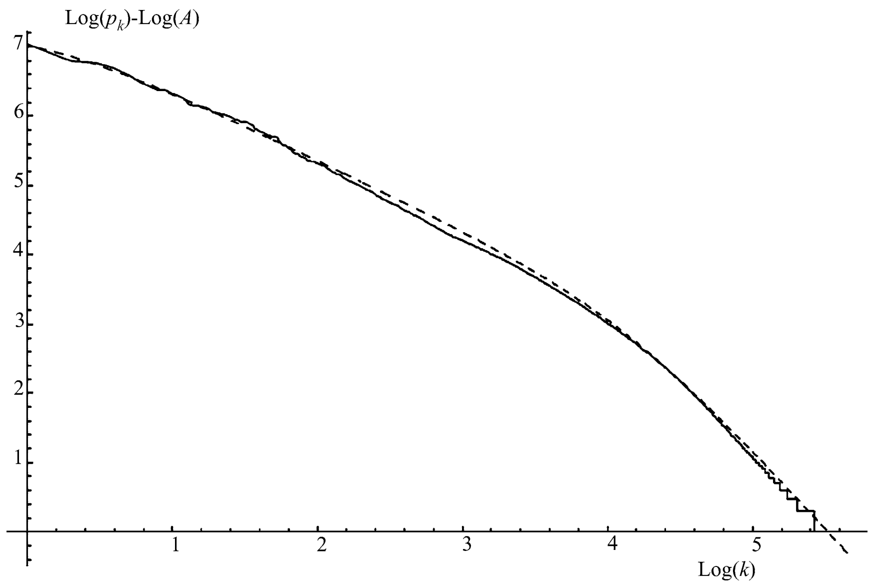

Fig. 9.3 Log–log plot of the frequency of words as a function of their ranking. Comparison of experimental data (solid line) and fitted curve (dotted line). Taken from [5]

$$f(r) \simeq \frac{A}{(1+Cr)^\alpha}. \tag{9.61}$$

Note that $\sum_r f(r) = N$, the total number of words in the text. Hence, the function $f(r)/N$ is an empirical probability distribution. In fact, it belongs to the q-exponential family with

$$\begin{aligned} c(r) &= 1, \\ q &= 1 + 1/\alpha, \\ \Phi(\theta) &= \frac{1}{C}\theta - \alpha, \\ H(r) &= r, \\ \theta &= \alpha C \left(\frac{N}{A}\right)^{1/\alpha}. \end{aligned} \tag{9.62}$$

If Zipf's law holds then the Figure 9.3 should show a straight line. This is only approximately the case. The experimental data show a cross-over from powerlaw decay with exponent about 1.05 to powerlaw decay with exponent about 2.3 — see [12] and the Figure 9.3. This requires a more complex modelling of the data, more complex than Zipf's law or Mandelbrot's modification. One possibility, considered here, is to rescale the data. In [5], a satisfactory fit of the 448,359 data points was obtained with the probability distribution

$$p(k) = \frac{A}{1 - \lambda + \lambda[1 + (1-\rho)\beta_0 k]^{1/(1-\rho)}}. \tag{9.63}$$

The exponent ρ was found to be $\rho = 0.568$. The expression can be written as (9.59), with

$$r = \rho, \Phi(\beta) = \frac{1}{A} - 1, \beta = \frac{\lambda}{A}, \text{ and } H(j) = j. \tag{9.64}$$

In other words, it maximises the Sharma-Mittal entropy with $r = \rho$ and $q = 0$.

Problems

9.1. Equivalent entropies
The two parameter Sharma-Mittal entropy functional was introduced in Section 9.7. It generalises Rényi's alpha entropy. Show that it also is a function of Tsallis' entropy functional. More precisely, show that a monotonically increasing function $\xi(u)$ exists such that $S_{q,r}(p) = \xi\left(S_q^{\text{Tsallis}}(p)\right)$.

Notes

Part of this Chapter follows the work of Marek Czachor and the author [5]. The calculation of the temperature of the configurational subsystem of a classical gas is taken from [3]. The Sections on fractals have been influenced by the work of Rudolf Riedi [13] and by the book of Christian Beck and of Friedrich Schlögl [4].

The notion of fractals and their importance for the description of nature have been pioneered by Benoit Mandelbrot [10]. The thermodynamic theory of dynamical systems was initiated in a series of mathematical papers by Ruelle, Bowen, Sinai. The multifractal analysis dates from the 1980s. In particular, the paper of Thomas Halsey et al [8] had a great influence.

The calculation of the generalised dimension of the asymmetric Cantor set, as presented here, is not the most elegant one. The partition of $[0,1]$ into sets of equal length 2^{-n} is not compatible with the arbitrary lengths l_1 and l_2 used during the construction. For that reason a two-parameter generating function was introduced in [8], see Section 11.4 of [4]. The notion of escort probability distributions is due to Beck and Schlögl [4].

Frank and Daffertshofer [6] and Frank and Plastino [7] introduced the Sharma-Mittal entropy in the physics literature. Marco Masi [11] relates the Sharma-Mittal entropy to what he calls the super-extensive entropy.

Objectives

- Explain the definition of non-linear Kolmogorov-Nagumo averages.
- Know about the definition of Rényi's alpha entropies by means of non-linear averages.
- Understand the construction of an asymmetric Cantor set.
- Know the definition of the generalised dimensions of a fractal measure with support in the interval $[0,1]$.
- Use the notion of escort probabilities to make a one-parameter family of probability distributions. Show that it belongs to the exponential family.
- Know about multifractal analysis. Explain how to derive a microcanonical description of a fractal measure.
- Explain the calculation of the generalised dimension of a Cantor set both using the canonical and the microcanonical description.
- Use Kolmogorov-Nagumo averages to show that the Sharma-Mittal entropy is a generalisation of Rényi's alpha entropies.
- Link the problem of optimising the Sharma-Mittal entropy to that of optimising the Tsallis entropy.
- Explain Zipf's law by means of an example.
- Relate Zipf's law to probability distributions belonging to the q-exponential family.

References

1. http://www.gutenberg.org/catalog/ 145
2. Selected Papers of Alfréd Rényi, Vol. 2. Akadémiai Kiadó, Budapest (1976) 132
3. Baeten, M., Naudts, J.: On the thermodynamics of classical microcanonical systems. arxiv:1009.1787 (2010) 66, 147
4. Beck, C., Schlögl, F.: Thermodynamics of chaotic systems: an introduction. Cambridge University Press (1997) 113, 147
5. Czachor, M., Naudts, J.: Thermostatistics based on Kolmogorov-Nagumo averages: Unifying framework for extensive and nonextensive generalizations. Phys. Lett. A **298**, 369–374 (2002) 143, 145, 146, 147
6. Frank, T., Daffertshofer, A.: Exact time-dependent solutions of the Rényi Fokker-Planck equation and the Fokker-Planck equations related to the entropies proposed by Sharma and Mittal. Physica A **285**, 351–366 (2000) 147
7. Frank, T., Plastino, A.: Generalized thermostatistics based on the Sharma-Mittal entropy and escort mean values. Eur. Phys. J. B **30**, 543–549 (2002) 147
8. Halsey, T.C., Jensen, M., Kadanoff, L., Procaccia, I., Shraiman, B.: Fractal measures and their singularities: The characterization of strange sets. Phys. Rev. A **33**, 1141–1151 (1986) 147
9. Mandelbrot, B.: Information theory and psycholinguistics: a theory of word frequencies. In: P. Lazarsfield, N. Henry (eds.) Readings in mathematical social sciences, pp. 151–168. MIT Press (1966) 145
10. Mandelbrot, B.: The fractal geometry of nature. Freeman, New York (1982) 147
11. Masi, M.: A step beyond Tsallis and Rényi entropies. Phys. Lett. A **338**, 217–224 (2005) 147
12. Montemurro, M.: Beyond the Zipf-Mandelbrot law in quantitative linguistics. Physica A **300**, 567–578 (2001) 145, 146
13. Riedi, R.: Multifractal processes. In: P. Doukhan, G. Oppenheim, M. Taqqu (eds.) Theory of Long-Range Dependence. Birkhäuser (2002) 141, 147
14. Sharma, B., Mittal, D.: New nonadditive measures of inaccuracy. J. Math. Sci. **10**, 28 (1975) 143
15. Tsallis, C., Bemski, G., Mendes, R.: Is re-association in folded proteins a case of nonextensivity? Phys. Lett. A **257**, 93–98 (1999) 144

Chapter 10
General Deformations

10.1 Deformed Exponential and Logarithmic Functions

What now follows is a further generalisation of the *q-deformed exponential and logarithmic functions,* discussed in Chapter 7.

Fix a strictly positive non-decreasing function $\phi(u)$, defined on the positive numbers $(0, +\infty)$. It is used to define a deformed logarithm by

$$\ln_\phi u = \int_1^u \mathrm{d}v \, \frac{1}{\phi(v)}. \tag{10.1}$$

Then $\ln_\phi(u)$ is a concave monotonically increasing function, satisfying $\ln_\phi(1) = 0$ and

$$\frac{\mathrm{d}}{\mathrm{d}u} \ln_\phi(u) = \frac{1}{\phi(u)}. \tag{10.2}$$

In particular, $\ln_\phi(u)$ is negative on $(0, 1)$ and positive on $(1, +\infty)$.

The natural logarithm is obtained with $\phi(u) = u$, the a-base logarithm $\log_a(u)$ is obtained with $\phi(u) = u \ln a$. The Tsallis q-logarithm (see (7.1)) is obtained with $\phi(u) = u^q$. The condition $q > 0$ is needed to make $\phi(u)$ increasing.

The inverse of the deformed logarithmic function is the deformed exponential function $\exp_\phi(u)$. It can be written with the help of some function $\psi(u)$ as

$$\exp_\phi(u) = 1 + \int_0^u \mathrm{d}v \, \psi(v). \tag{10.3}$$

By writing it in this way it is clear that $\exp_\phi(0) = 1$ and

$$\frac{\mathrm{d}}{\mathrm{d}u} \exp_\phi(u) = \psi(u). \tag{10.4}$$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_10,

The function $\psi(u)$ can be calculated once the function $\phi(u)$ is given. Indeed, by taking the derivative of

$$u = \exp_\phi(\ln_\phi(u)) \tag{10.5}$$

one obtains

$$1 = \psi(\ln_\phi(u))\frac{1}{\phi(u)}. \tag{10.6}$$

This can be written as

$$\phi(u) = \psi(\ln_\phi(u)). \tag{10.7}$$

In other words, $v = \ln_\phi(u)$ implies $\psi(v) = \phi(u)$. If v is smaller than $\ln_\phi(u)$ for all u then one can define $\psi(v) = 0$. If v is larger than $\ln_\phi(u)$ for all u then one can define $\psi(v) = +\infty$.

Note that $\psi(v)$ cannot be negative and is an increasing function. From (10.4) then follows that $\exp_\phi(u)$ is an increasing convex function of u. For further use, note also that

$$\psi(u) = \frac{\mathrm{d}}{\mathrm{d}u}\exp_\phi(u) = \phi\left(\exp_\phi(u)\right). \tag{10.8}$$

This relation generalises the well-known property that the derivative of the natural exponential is the exponential function itself.

10.2 Dual Definitions

Let us define dual deformed functions by

$$\exp^*_\phi(u) = \frac{1}{\exp_\phi(-u)} \tag{10.9}$$

and

$$\ln^*_\phi(u) = -\ln_\phi(1/u). \tag{10.10}$$

These dual functions not always satisfy the requirements for being again a deformed exponential or deformed logarithm. Indeed, from

$$\frac{\mathrm{d}}{\mathrm{d}u}\ln^*_\phi(u) = \frac{1}{u^2\phi(1/u)} \tag{10.11}$$

follows the requirement that the function

$$\phi^*(u) = u^2\phi(1/u) \tag{10.12}$$

should be an increasing function of u. From

$$\frac{\mathrm{d}}{\mathrm{d}u}\exp^*_\phi(u) = \left(\exp^*_\phi(u)\right)^2 \psi(-u) \tag{10.13}$$

then follows that

$$\exp^*_\phi(u) = 1 + \int_0^u \mathrm{d}v\,\psi^*(v), \tag{10.14}$$

with

$$\psi^*(u) = \left(\exp^*_\phi(u)\right)^2 \psi(-u). \tag{10.15}$$

In the case of the natural logarithm is $\phi^*(u) = \phi(u) = u$ and $\psi^*(u) = \psi(u) = e^u$. Therefore, the natural logarithm is said to be *self-dual*. However, there exist non-trivial examples of self-dual deformed exponential and logarithmic functions — See the Box 10.1.

10.3 Deduced Logarithms

When using a deformed logarithm to define entropy one notices that the following property of the natural logarithm gets lost

$$\frac{\mathrm{d}}{\mathrm{d}u}\left(u\ln\frac{1}{u}\right) = -\ln u - 1. \tag{10.22}$$

This is the reason to introduce the notion of *deduced logarithm* $\omega_\phi(u)$. It requires the additional condition that the possible divergence of $\ln_\phi(u)$ at $u = 0$ is weak enough so that the integral

$$-\int_0^1 \mathrm{d}u\,\ln_\phi(u) = \int_0^1 \mathrm{d}v\,\frac{v}{\phi(v)} < +\infty \tag{10.23}$$

converges (use partial integration to obtain the above relation).

The deduced logarithm is defined by

$$\omega_\phi(u) = u\int_0^{1/u} \mathrm{d}v\,\frac{v}{\phi(v)} - \int_0^1 \mathrm{d}v\,\frac{v}{\phi(v)} - \ln_\phi\frac{1}{u}. \tag{10.24}$$

One verifies immediately that $\omega_\phi(1) = 0$ and that

$$\frac{\mathrm{d}}{\mathrm{d}u}u\omega_\phi\left(\frac{1}{u}\right) = -\ln_\phi(u) - \int_0^1 \mathrm{d}v\,\frac{v}{\phi(v)}. \tag{10.25}$$

An example of *self-dual* deformed exponential and logarithmic functions is found in the work of Kaniadakis [2, 3]. Let $-1 < \kappa < 1$, and define

$$\exp_\kappa(u) = \left(\kappa u + \sqrt{1+\kappa^2 u^2}\right)^{1/\kappa}. \tag{10.16}$$

It is strictly positive and finite for all real u. The inverse function is

$$\ln_\kappa(u) = \frac{1}{2\kappa}\left(u^\kappa - u^{-\kappa}\right). \tag{10.17}$$

In the limit $\kappa = 0$ these functions coincide with the usual definitions of logarithmic and exponential functions. By taking the derivative of (10.17) one finds that $\ln_\kappa(u) = \ln_\phi(u)$ with $\phi(u)$ given by

$$\phi(u) = \frac{2u}{u^\kappa + u^{-\kappa}}. \tag{10.18}$$

The derivative of this function is

$$\phi'(u) = 2\frac{(1-\kappa)u^\kappa + (1+\kappa)u^{-\kappa}}{(u^\kappa + u^{-\kappa})^2}. \tag{10.19}$$

Hence $\phi(u)$ is a positive increasing function, given $-1 \le \kappa \le 1$. Therefore, $\ln_\kappa(u)$ is a deformed logarithm. One verifies that

$$\begin{aligned}\exp_\kappa(-u)\exp_\kappa(u) &= \left(-\kappa u + \sqrt{1+\kappa^2u^2}\right)^{1/\kappa}\left(\kappa u + \sqrt{1+\kappa^2u^2}\right)^{1/\kappa}\\ &= 1,\end{aligned} \tag{10.20}$$

The latter implies that the functions are self-dual.

Next calculate $\chi(u)$ using (10.27)

$$\chi(u) = \frac{2}{\int_0^{1/u} \mathrm{d}v\,(v^\kappa + v^{-\kappa})} = \frac{2(1-\kappa^2)u^2}{(1+\kappa)u^{1+\kappa} + (1-\kappa)u^{1-\kappa}}. \tag{10.21}$$

One verifies that $\frac{\mathrm{d}}{\mathrm{d}u}\chi(u) = \frac{u^\kappa + u^{-\kappa}}{2u^2}\chi(u)^2 \ge 0$. Hence, the logarithm deduced from $\ln_\kappa(u)$ is $\ln_\chi(u)$ with $\chi(u)$ given by (10.21).

Box 10.1 Example of self-dual deformed functions

In fact, $\omega_\phi(u)$ is again a deformed logarithm — see the Box 10.2.

The relation (10.25) generalises (10.22). In the case of the natural logarithm is $\phi(u) = u$ so that $\omega_\phi(u) = \ln u$. In the case of a q-deformed logarithm is $\phi(u) = u^q$ so that (using the notation ω_q instead of ω_ϕ)

$$\omega_q(u) = -\frac{1}{2-q}\ln_q\frac{1}{u} = \frac{1}{2-q}\ln_q^* u = \frac{1}{2-q}\ln_{2-q} u. \tag{10.26}$$

Here we prove that the deduced logarithm $\omega_\phi(u)$ is again a deformed logarithm. This means that there exists a increasing positive function $\chi(u)$ such that $\omega_\phi(u) = \ln_\chi(u)$. Let $\chi(u)$ defined by

$$\chi(u) = \frac{1}{\int_0^{1/u} \mathrm{d}v \, \frac{v}{\phi(v)}}. \tag{10.27}$$

Then one verifies that

$$\frac{\mathrm{d}}{\mathrm{d}u}\omega_\phi(u) = \frac{1}{\chi(u)} \tag{10.28}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}u}\chi(u) = \frac{\chi(u)^2}{u^3\phi(1/u)} > 0. \tag{10.29}$$

Hence, one concludes that $\omega_\phi(u) = \ln_\chi(u)$ is a deformed logarithm.

Box 10.2 Proof that the deduced logarithm $\omega_\phi(u)$ is again a deformed logarithm

10.4 The Phi-Exponential Family

Given a strictly positive monotonically increasing function $\phi(u)$ the family of probability distributions $p_\theta(x)$ is said to belong to the *ϕ-exponential family* if it can be written into the form

$$p_\theta(x) = c(x) \exp_\phi\left(-\alpha(\theta) - \theta^k H_k(x)\right). \tag{10.30}$$

The function $\alpha(\theta)$ is needed to guarantee the normalisation of the probability distribution. Its physical meaning is that of a Massieu function. However, in general it does not coincide with the Massieu function $\Phi(\theta)$. One can show that it is always a convex function of the parameters θ — see the Box 10.3.

If the function $\phi(u)$ is linear then the definition coincides with the standard definition of the exponential family. If $\phi(u) = u^q$ then it coincides with the definition of the q-exponential family, given in Chapter 7.

The function $\phi(u)$ in (10.30) may be stochastic. This means that it may also depend on the variable x. An example where this is needed follows in Section 10.7.

10.5 Escort Probabilities

Inspired by the definition in the case of the q-exponential family it is obvious to define the *escort probability distribution* $P_\theta(x)$ of a probability distribution

Here we prove that in the case of the ϕ-exponential family the function $\alpha(\theta)$ is always convex.

Fix two sets op parameters θ and η and a number λ between 0 and 1. Because the deformed exponential function $\exp_\phi(u)$ is convex one has

$$\begin{aligned} & c(x)\exp_\phi\left[-\lambda\alpha(\theta)-(1-\lambda)\alpha(\eta)-(\lambda\theta^k+(1-\lambda\eta^k))H_k(x)\right] \\ \leq\; & \lambda c(x)\exp_\phi[-\alpha(\theta)-\theta^k H_k(x)]+(1-\lambda)c(x)\exp_\phi[-\alpha(\eta)-\eta^k H_k(x)] \\ =\; & \lambda p_\theta(x)+(1-\lambda)p_\eta(x). \end{aligned} \tag{10.31}$$

Integrating this expression over the phase space gives

$$\int \mathrm{d}x\, c(x)\exp_\phi\left[-\lambda\alpha(\theta)-(1-\lambda)\alpha(\eta)-(\lambda\theta^k+(1-\lambda\eta^k)H_k(x)\right] \leq 1. \tag{10.32}$$

On the other hand one has

$$\begin{aligned} 1 &= \int \mathrm{d}x\, p_{\lambda\theta+(1-\lambda)\eta}(x) \\ &= \int \mathrm{d}x\, c(x)\exp_\phi\left[-\alpha(\lambda\theta+(1-\lambda)\eta)-(\lambda\theta^k+(1-\lambda\eta^k)H_k(x)\right]. \end{aligned} \tag{10.33}$$

Comparing (10.32) with (10.32), and using that $\exp_\phi(u)$ is an increasing function, one concludes that

$$-\lambda\alpha(\theta)-(1-\lambda)\alpha(\eta)\leq -\alpha(\lambda\theta+(1-\lambda)\eta). \tag{10.34}$$

This means that $\alpha(\theta)$ is a convex function.

Box 10.3 Proof of convexity of the function $\alpha(\theta)$

$p_\theta(x)$, belonging to the ϕ-exponential family, by

$$P_\theta(x)=\frac{c(x)}{z(\theta)}\phi\left(\frac{p_\theta(x)}{c(x)}\right). \tag{10.35}$$

Here, $z(\theta)$ is the normalisation

$$z(\theta)=\int \mathrm{d}x\, c(x)\phi\left(\frac{p_\theta(x)}{c(x)}\right), \tag{10.36}$$

assuming that this integral is convergent. Using $\phi(\exp_\phi(u))=\psi(u)$ and the definition of $p_\theta(x)$ one obtains

$$z(\theta)=\int \mathrm{d}x\, c(x)\psi\left(-\alpha(\theta)-\theta^k H_k(x)\right). \tag{10.37}$$

Clearly, in the linear case $\phi(u) = u$ the escort probability $P_\theta(x)$ coincides with $p_\theta(x)$ and the normalisation $z(\theta)$ is identically 1. If $\phi(u) = u^q$ then the definition (7.24) is recovered.

The derivatives of the probability $p_\theta(x)$ with respect to the parameters θ^k are proportional to the escort probability. To see this, use (10.8) to calculate

$$\begin{aligned}\frac{\partial}{\partial\theta^k} p_\theta(x) &= c(x)\phi\left(\frac{p_\theta(x)}{c(x)}\right)\left[-\frac{\partial\alpha}{\partial\theta^k} - H_k(x)\right]\\ &= z(\theta)P_\theta(x)\left[-\frac{\partial\alpha}{\partial\theta^k} - H_k(x)\right].\end{aligned} \tag{10.38}$$

This expression can be used to find out whether a probability distribution belongs to the ϕ-exponential family and what the correct choice is of the function $\phi(u)$ — See the next Section.

Integrating (10.38) gives

$$\begin{aligned}0 &= \frac{\partial}{\partial\theta^k}\int \mathrm{d}x\, p_\theta(x)\\ &= z(\theta)\left[-\frac{\partial\alpha}{\partial\theta^k} - \langle\langle H_k\rangle\rangle_\theta\right].\end{aligned} \tag{10.39}$$

As before, $\langle\langle H_k\rangle\rangle_\theta$ denotes the average of $H_k(x)$ with respect to the escort probability $P_\theta(x)$. Expression (10.39) shows that minus the derivative of the function $\alpha(\theta)$ equals the expectation of the estimator $H_k(x)$ with respect to the escort probability distribution. Hence, averages with respect to the escort probability can be used to estimate the parameters θ^k — see Chapter 2, equation (2.5).

A more explicit expression for the escort probability $P_\theta(x)$ involves the function $\psi(x)$ associated with $\phi(x)$ by 10.8. From the definition (10.35) follows

$$\begin{aligned}P_\theta(x) &= \frac{c(x)}{z(\theta)}\phi\left(\exp_\phi[-\alpha(\theta) - \theta^k H_k]\right)\\ &= \frac{c(x)}{z(\theta)}\psi[-\alpha(\theta) - \theta^k H_k].\end{aligned} \tag{10.40}$$

This implies the following expression for the normalisation $z(\theta)$

$$z(\theta) = \int \mathrm{d}x\, c(x)\psi[-\alpha(\theta) - \theta^k H_k]. \tag{10.41}$$

10.6 Method to Test Phi-Exponentiality

Expression (10.38) can be used to determine whether a given probability distribution belongs to the ϕ-exponential family. The method, described below, will be demonstrated for the *Cauchy distribution*, which is known not to belong to the exponential family — See (2.17).

1) Calculate the derivatives of the probability distribution.

From

$$p_\omega(x) = \frac{1}{\pi}\frac{\omega}{x^2+\omega^2} \tag{10.42}$$

follows

$$\frac{\mathrm{d}}{\mathrm{d}\omega}p_\omega(x) = p_\omega(x)\left[\frac{1}{\omega} - \frac{2\omega}{x^2+\omega^2}\right]. \tag{10.43}$$

2) Try to separate the x-dependence from the parameter dependence. Use the freedom to introduce an escort probability distribution and, if needed, new parameters.

The first term within the square brackets of (10.43) depends only on ω, not on the variable x. Hence it is in the right form. However, the second term still depends on both ω and x. Write (10.43) therefore into the form

$$\frac{\mathrm{d}}{\mathrm{d}\omega}p_\omega(x) = \frac{\pi}{\omega^2}p_\omega(x)^2\left[-\omega^2+x^2\right]. \tag{10.44}$$

Now the separation within the square brackets is accomplished. The factor in front of the square brackets must be the escort probability distribution, multiplied by its normalisation. Choose therefore $\phi(x) = x^2$. The corresponding deformed logarithm and exponential functions are

$$\begin{aligned}
\ln_2(u) &= 1 - \frac{1}{u} && u > 0,\\
\exp_2(u) &= \frac{1}{1-u} && u < 1,\\
&= +\infty && \text{otherwise.}
\end{aligned} \tag{10.45}$$

The escort probability is then

$$P_\omega(x) = \frac{1}{z(\omega)}p_\omega(x)^2 \tag{10.46}$$

with

$$z(\omega) = \frac{\omega^2}{\pi^2}\int \mathrm{d}x\,\frac{1}{(x^2+\omega^2)^2}. \tag{10.47}$$

Expression (10.44) then becomes

$$\frac{\mathrm{d}}{\mathrm{d}\omega}p_\omega(x) = \frac{\pi}{\omega^2}z(\omega)P_\omega(x)\left[-\omega^2+x^2\right]. \tag{10.48}$$

This is almost (10.38). Introduce the parameter $\theta = -\pi/\omega$. Then one obtains (10.38) with $H(x) = -x^2$ and $\partial\alpha/\partial\theta = \omega^2$.

The final step of the method is the verification that $p_\theta(x)$ belongs indeed to the ϕ-exponential family. This verification is necessary because integration of (10.38) may produce integration constants which depend on the variables x. If this is the case then $p_\theta(x)$ is not of the form, desired to belong to the ϕ-exponential family. Then one can try to introduce extra parameters θ^k in such a way that the integration constants become additional Hamiltonians H_k.

In the case of the Cauchy distribution the verification is straightforward. Indeed, equating

$$\exp_2(-\Phi(\theta) + \theta x^2) = \frac{1}{1 + \Phi(\theta) - \theta x^2} \tag{10.49}$$

to (10.42) yields the condition

$$1 + \Phi(\theta) - \theta x^2 = \frac{\pi}{\omega} x^2 + \pi\omega. \tag{10.50}$$

This is satisfied with $\theta = -\pi/\omega$ and $\Phi(\theta) = \pi\omega - 1 = -\pi^2/\theta - 1$. This shows that the Cauchy distribution belongs to the curved ϕ-exponential family with $\phi(u) = u^2$.

10.7 Example: The Site Percolation Problem

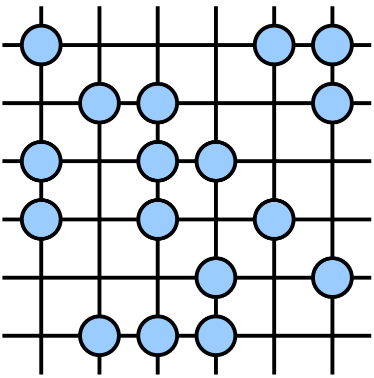

Fig. 10.1 Occupied sites of a square lattice

Consider a lattice, for instance the square lattice. Colour the sites of the lattice black with probability p, white with probability $1 - p$ in an *iid* manner. See the Figure 10.1. The origin of the lattice belongs to a cluster of s black sites with a certain probability which depends on p, or is white with probability $1-p$. If the probability p is large enough then there exists a non-vanishing probability $p(\infty)$ that the origin belongs to an infinite black cluster. Then the lattice is said to *percolate*. See the Figure 10.2. The critical probability p_c is the lower limit of p-values for which $p(\infty) \neq 0$. For a square lattice numerical simulations show that $p_c \simeq 0.593$.

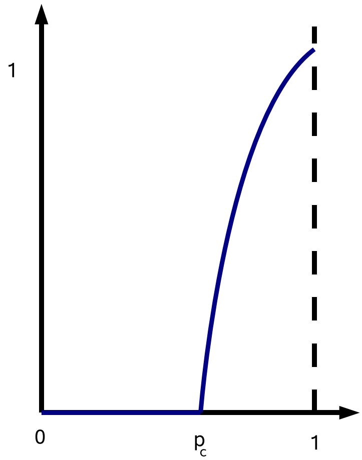

Fig. 10.2 Sketch of the probability of percolation as a function of the probability p of occupying a lattice site

Two clusters with the same number of black sites n can have a different shape. These shapes are called *lattice animals*. All possible shapes will be numbered $j = 1, 2, 3, \cdots$. The label $j = 0$ is reserved for the empty cluster. The number of clusters with the same shape is denoted $c(j)$. The number of sites in clusters of shape j is denoted $s(j)$. Some of the sites of the shape are internal, others are at the perimeter. The number of white sites that touch the cluster at one of its perimeter sites is denoted $t(j)$. The probability that the origin belongs to a cluster of shape j is then given by

$$p(j) = c(j)p^{s(j)}(1-p)^{t(j)}. \tag{10.51}$$

This result is also valid for $j = 0$ if we adopt the conventions that $c(0) = 1$, $s(0) = 0$, and $t(0) = 1$.

The normalisation condition can be written as

$$p(\infty) + \sum_{j=0}^{\infty} p(j) = 1. \tag{10.52}$$

This expression can be used to calculate the percolation probability $p(\infty)$.

Let us now try to write (10.51) in the form of an exponential family. One has

$$\begin{aligned} p(j) &= c(j)\exp\left[s(j)\ln p + t(j)\ln(1-p)\right] \\ &= c(j)\exp\left[\left(\ln p - \frac{t(j)}{s(j)+t(j)}\ln\frac{p}{1-p}\right)(s(j)+t(j))\right] \\ &= c(j)\exp\left[\left(-\alpha(\beta) - \beta H(j)\right)(s(j)+t(j))\right], \end{aligned} \tag{10.53}$$

with

$$\begin{aligned} \beta &= \ln\frac{p}{1-p} \\ H(j) &= \frac{t(j)}{s(j)+t(j)} \\ \alpha(\beta) &= -\ln p. \end{aligned} \tag{10.54}$$

Introduce now a stochastic function $\phi(u)$ defined by

$$\phi_j(u) = (s(j)+t(j))u. \tag{10.55}$$

Then $\exp_{\phi_j}(v) = \exp[(s(j)+t(j))v]$ so that

$$p(j) = c(j)\exp_{\phi_j}\left[-\alpha(\beta) - \beta H(j)\right]. \tag{10.56}$$

Assume now for simplicity that $p < p_c$ so that $p(\infty) = 0$. Then the $p(j)$ sum up to 1 and form a probability distribution belonging to the ϕ-exponential family.

The escort probability distribution is given by

$$\begin{aligned} P(j) &= \frac{1}{z(\beta)}\phi_j\left(p(j)\right) \\ &= \frac{1}{z(\beta)}p(j)(s(j)+t(j)), \end{aligned} \tag{10.57}$$

with

$$\begin{aligned} z(\beta) &= \sum_{j=0}^{\infty} p(j)(s(j)+t(j)) \\ &= \langle s\rangle_\beta + \langle t\rangle_\beta. \end{aligned} \tag{10.58}$$

From expression (10.54) follows

$$\begin{aligned} \langle\langle H\rangle\rangle &= -\frac{\mathrm{d}\alpha}{\mathrm{d}\beta} \\ &= 1-p. \end{aligned} \tag{10.59}$$

On the other hand is

$$\begin{aligned}\langle\langle H\rangle\rangle &= \frac{1}{z(\beta)}\langle (s+t)\frac{t}{s+t}\rangle_\beta\\ &= \frac{\langle t\rangle_\beta}{\langle s\rangle_\beta + \langle t\rangle_\beta}.\end{aligned} \tag{10.60}$$

One has therefore the identity

$$\langle t\rangle_\beta = (1-p)\left(\langle s\rangle_\beta + \langle t\rangle_\beta\right), \qquad p < p_c. \tag{10.61}$$

10.8 Generalised Quantum Statistics

Fix a strictly positive non-decreasing function $\phi(u)$[1]. The quantum model with density operators ρ_θ belongs to the *quantum ϕ-exponential family* if there exist self-adjoint operators H_k such that

$$\rho_\theta = \exp_\phi(-\alpha(\theta) - \theta^k H_k). \tag{10.62}$$

The function $\alpha(\theta)$ is used for normalisation. In the case $\phi(u) = u^q$ it is possible to take the normalisation in front of the deformed exponential — see the Box 10.4

In the case that $\phi(u) = u^q$ then (10.62) yields

$$\begin{aligned}\rho_\theta &= \exp_q\left(-\alpha(\theta) - \theta^k H_k\right)\\ &= \left[1-(1-q)\alpha(\theta) - (1-q)\theta^k H_k\right]_+^{1/(1-q)}\\ &= \frac{1}{Z(\eta)}\left[1-(1-q)\eta^k H_k\right]_+^{1/(1-q)}\\ &= \frac{1}{Z(\eta)}\exp_q(-\eta^k H_k),\end{aligned} \tag{10.63}$$

with η^k and $Z(\eta)$ as in Section 7.3 on the quantum exponential family, equations (7.14, 7.15). The normalisation $Z(\eta)$ can be calculated from

$$Z(\eta) = \operatorname{Tr}\exp_q(-\eta^k H_k). \tag{10.64}$$

Box 10.4 The deformed quantum exponential family in the case $\phi(u) = u^q$

It is obvious to define *escort density operators* σ_θ by

$$\sigma_\theta = \frac{1}{z(\theta)}\phi\left(\rho_\theta\right) \tag{10.65}$$

[1] In the quantum case the function $\phi(u)$ must not be stochastic.

with

$$z(\theta) = \operatorname{Tr} \phi\left(\rho_\theta\right). \tag{10.66}$$

The basic property of the escort density operators is their relation with the θ^k-derivatives of ρ_θ. However, as before, in Section 2.7, the operators H_k do not necessarily commute between each other. Hence, it is not straightforward to calculate the derivatives of ρ_θ. But using cyclic permutation under the trace and (10.8) one calculates

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta^k} \operatorname{Tr} \rho_\theta \\ &= \operatorname{Tr} \frac{\partial}{\partial\theta^k} \exp_\phi(-\alpha(\theta) - \theta^k H_k) \\ &= \operatorname{Tr} \phi(\rho_\theta) \frac{\partial}{\partial\theta^k} \left[-\alpha(\theta) - \theta^l H_l\right] \\ &= z(\theta) \left[-\frac{\partial\alpha}{\partial\theta^k} - \langle\langle H_k \rangle\rangle_\theta\right], \end{aligned} \tag{10.67}$$

with

$$\langle\langle H_k \rangle\rangle_\theta = \operatorname{Tr} \sigma_\theta H_k. \tag{10.68}$$

This implies

$$-\frac{\partial\alpha}{\partial\theta^k} = \langle\langle H_k \rangle\rangle_\theta. \tag{10.69}$$

Hence, the quantum expectation of the estimators H_k with respect to the escort density operator can be used to estimate the parameters θ^k.

Problems

10.1. Heine's distribution
See for instance [1]. Fix $0 < q < 1$. Then Heine's probability distribution is given by

$$p_\lambda(n) = \lambda^n e_q(-\lambda) \frac{q^{n(n-1)/2}}{[n]_q!}, \quad n = 0, 1, 2, \cdots, \lambda > 0, \tag{10.70}$$

with $e_q(-\lambda)$ the *q-deformed exponential function* introduced by *Jackson*

$$e_q(-\lambda) = \prod_{j=1}^{\infty} \frac{1}{1 + \lambda(1-q)q^{j-1}}, \tag{10.71}$$

and $[n]_q!$ the q-deformed factorial given by

$$[n]_q! = [1]_q[2]_q \cdots [n_q] \quad \text{with} \quad [n_q] = 1 + q + \cdots + q^{n-1}. \qquad (10.72)$$

Show that $p_\lambda(n)$ is a special 1-parameter case of a 2-parameter exponential family.

10.2. Lambert's W function
Let $W(u)$ be the solution of $u = W(u)\exp(W(u))$ which satisfies $W(u) \geq -1$ on the interval $-\frac{1}{e} \leq u$. Introduce a deformed logarithm defined by $\ln_W(u) = W((u-1)/e)$. Verify its properties. Verify that the corresponding deformed exponential function is given by $\exp_W(v) = 1 + v\exp(v+1)$ if $v \geq -1$ and by $\exp(v) = 0$ for $v \leq -1$.

Notes

The generalisation presented in this Chapter was developed by the author in a series of papers [4, 6, 7, 10, 5, 8]. The percolation problem was treated in [9].

A general reference concerning the percolation problem is the book of Dieter Stauffer [11].

Objectives

- Know the definitions and properties of ϕ-deformed logarithmic and exponential functions, where $\phi(u)$ is an arbitrary positive non-decreasing function on the positive axis.
- Know the definition of the ϕ-exponential family, and of the corresponding escort probability distributions.
- Be able to test whether a parametrised family of probability distributions belongs to a ϕ-exponential family.
- Give an example of a parametrised family of probability distributions with a non-trivial escort family.
- Know about the percolation problem.
- Know about the quantum ϕ-exponential family.

References

1. Charalambides, C.A., Papadatos, N.: The q-factorial moments of discrete q-distributions and a characterization of the Euler distribution. In: N. Balakr-

ishnan, I.G. Bairamov, O.L. Gebizlioglu (eds.) Advances on Models, Characterizations and Applications, pp. 57–71. Chapman & Hall/CRC Press, Boca Raton (2005) 161
2. Kaniadakis, G.: Non-linear kinetics underlying generalized statistics. Physica A **296**, 405–425 (2001) 152
3. Kaniadakis, G., Scarfone, A.: A new one parameter deformation of the exponential function. Physica A **305**, 69–75 (2002) 128, 152, 176
4. Naudts, J.: Deformed exponentials and logarithms in generalized thermostatistics. Physica A **316**, 323–334 (2002) 112, 162
5. Naudts, J.: Estimators, escort probabilities, and phi-exponential families in statistical physics. J. Ineq. Pure Appl. Math. **5**, 102 (2004) 112, 162, 177
6. Naudts, J.: Generalized thermostatistics and mean field theory. Physica A **332**, 279–300 (2004) 129, 162
7. Naudts, J.: Generalized thermostatistics based on deformed exponential and logarithmic functions. Physica A **340**, 32–40 (2004) 162
8. Naudts, J.: Escort operators and generalized quantum information measures. Open Systems and Information Dynamics **12**, 13–22 (2005) 162, 177
9. Naudts, J.: Parameter estimation in nonextensive thermostatistics. Physica A **365**, 42–49 (2006) 36, 162
10. Naudts, J.: Continuity of a class of entropies and relative entropies. Rev. Math. Phys. **16**, 809822 (2004); Errata. Rev. Math. Phys. **21**, 947–948 (2009) 129, 162, 177
11. Stauffer, D.: Introduction to percolation theory. Plenum Press, New York (1985) 162

Chapter 11
General Entropies

11.1 The Entropy Functional

The definition of deformed entropy needed in the present context is the one which most naturally adapts to the definition of the ϕ-exponential family in the previous Chapter. We have seen in Section 8.2 that the Tsallis entropy $S_q^{\mathrm{Tsallis}}(p)$ is not associated in a natural way with the probability distributions of the q-exponential family. It is desirable that a generalised entropy $S_\phi(p)$ is maximised by probability distributions belonging to the ϕ-exponential family. For that reason, it is defined below using the *deduced logarithmic function* $\omega_\phi(u)$ — see (10.24) for the definition of this concept.

The *ϕ-deformed entropy functional* is defined by

$$\begin{aligned} S_\phi(p) &= s\sum_j p(j)\omega_\phi\left(\frac{1}{p(j)}\right) \\ &= s\sum_j \int_0^{p(j)} \mathrm{d}v\,\frac{v}{\phi(v)} - s\sum_j p(j)\ln_\phi(p(j)) - 1 \\ &= -s\sum_j \int_0^{p(j)} \mathrm{d}v\,\ln_\phi(v) - 1, \end{aligned} \tag{11.1}$$

with

$$\frac{1}{s} = \int_1^0 \mathrm{d}v\,\ln_\phi(v) = \int_0^1 \mathrm{d}v\,\frac{v}{\phi(v)}. \tag{11.2}$$

The last line of (11.1) is obtained by partial integration. Throughout this Chapter it is assumed that the integral in (11.2) converges. This is the condition for the deduced logarithm $\omega_\phi(u)$ to exist, but is for instance not satisfied when $\phi(u) = u^2$.

By definition, $S_\phi(p)$ is a sum of non-negative terms. Hence, it either converges or it diverges towards $+\infty$

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_11,

$$0 \le S_\phi(p) \le +\infty. \tag{11.3}$$

It is a concave function of p — see the Box 11.1.

In the continuous case, the definition becomes

$$\begin{aligned} S_\phi(p) &= s \int \mathrm{d}x\, p(x) \omega_\phi \left(\frac{c(x)}{p(x)} \right) \\ &= s \int \mathrm{d}x\, c(x) \int_0^{p(x)/c(x)} \mathrm{d}v\, \frac{v}{\phi(v)} \\ &\quad - s \int \mathrm{d}x\, p(x) \ln_\phi \left(\frac{p(x)}{c(x)} \right) - 1 \\ &= -s \int \mathrm{d}x\, c(x) \int_0^{p(x)/c(x)} \mathrm{d}v\, \ln_\phi(v) - 1. \end{aligned} \tag{11.4}$$

It still is a concave function of the probability distribution p. However, it is not necessarily positive because $p(x)/c(x)$ can exceed 1, resulting in negative contributions to the integral in the first line of (11.4).

Concavity of $S_\phi(p)$ means that

$$S_\phi(\lambda p + (1-\lambda) q) \ge \lambda S_\phi(p) + (1-\lambda) S_\phi(q) \tag{11.5}$$

for all λ in $[0, 1]$. This follows from the last line of (11.1). Indeed, $\ln_\phi(v)$ is an increasing function of v. Hence the second derivative of

$$f(u) = \int_0^u \mathrm{d}v\, \ln_\phi(v) \tag{11.6}$$

cannot be negative. This means that $f(u)$ is convex, so that

$$\begin{aligned} &\int_0^{\lambda p(j) + (1-\lambda) q(j)} \mathrm{d}v\, \ln_\phi(v) \\ &\le \lambda \int_0^{p(j)} \mathrm{d}v\, \ln_\phi(v) + (1-\lambda) \int_0^{q(j)} \mathrm{d}v\, \ln_\phi(v). \end{aligned} \tag{11.7}$$

In combination with (11.1) this implies (11.5).

Box 11.1 Concavity of the ϕ-deformed entropy functional

11.2 Relative Entropies

The relative entropy (or divergence) used below is the *relative entropy of the Bregman type* — see (8.47).

The function $f(u)$ in the definition (8.47) of the relative entropy is taken equal to

$$f(u) = s\int_1^u \mathrm{d}v\ \ln_\phi(v) + 1 - u. \tag{11.8}$$

Note that this function is convex. Its second derivative equals $f''(u) = s/\phi(u)$. The result is

$$\begin{aligned} D_\phi(p||q) &= s\sum_j \left[f(p(j)) - f(q(j)) - (p(j) - q(j))f'(q(j))\right] \\ &= s\sum_j \int_{q(j)}^{p(j)} \mathrm{d}v\ \ln_\phi(v) - s\sum_j [p(j) - q(j)]\ln_\phi(q(j)) \\ &= s\sum_j \int_{q(j)}^{p(j)} \mathrm{d}v\ [\ln_\phi(v) - \ln_\phi(q(j))] \\ &\leq +\infty. \end{aligned} \tag{11.9}$$

Positivity $D_\phi(p||q) \geq 0$ follows because $\ln_\phi(v)$ is an increasing function. Indeed, if $p(j) > q(j)$ then $v \geq q(j)$ and hence $\ln_\phi(v) - \ln_\phi(q(j)) \geq 0$. Hence, the contribution to the last line of (11.9) is positive. On the other hand, if $p(j) < q(j)$ then the argument of the integral is negative. But the integration is in the negative sense. Hence, also in this case the contribution is positive. One concludes that all terms of (11.9) are positive.

Note further that $D_\phi(p||p) = 0$ and that $D_\phi(p||q) = 0$ implies that $p = q$. Finally, $D_\phi(p||q)$ is convex in the first argument. This follows immediately from

$$D_\phi(p||q) = S_\phi(q) - S_\phi(p) - s\sum_j [p(j) - q(j)]\ln_\phi(q(j)) \tag{11.10}$$

because $S_\phi(p)$ is concave.

11.3 The Variational Principle

The relative entropy $D_\phi(p||q)$ can be used to prove the *variational principle*, stating that

$$S_\phi(p) - s\theta^k\langle H_k\rangle_p \tag{11.11}$$

is maximal if and only if the probability distribution p equals p_θ as given by (10.30). See the Box 11.2 for a proof of this statement. This characterizes the generalized exponential family as the family of pdfs which maximize (11.11).

From (11.10) and the definition of p_θ follows

$$\begin{aligned} D_\phi(p||p_\theta) &= S_\phi(p_\theta) - S_\phi(p) - s\sum_j [p(j) - p_\theta(j)] \ln_\phi(p_\theta(j)) \\ &= S_\phi(p_\theta) - S_\phi(p) - s\sum_j [p(j) - p_\theta(j)] \left[-\alpha(\theta) - \theta^k H_k(j)\right] \\ &= S_\phi(p_\theta) - S_\phi(p) + s\theta^k[\langle H_k\rangle_p - \langle H_k\rangle_\theta]. \end{aligned} \tag{11.12}$$

Because $D_\phi(p||p_\theta) \geq 0$, this shows that for all p

$$S_\phi(p_\theta) - s\theta^k\langle H_k\rangle_\theta \geq S_\phi(p) - s\theta^k\langle H_k\rangle_p. \tag{11.13}$$

Moreover, the equality holds if and only if $p = p_\theta$.

Box 11.2 Proof of the variational principle

By definition, the *Massieu function* $\Phi(\theta)$ is the Legendre transform of the entropy $S_\phi(p_\theta)$. But note that we use $s\theta^k$ as the transformation variables instead of the θ^k. Hence Massieu's function is equal to the maximal value attained by (11.11), this is,

$$\Phi(\theta) = S_\phi(p_\theta) - s\theta^k\langle H_k\rangle_\theta \tag{11.14}$$

and

$$\frac{\partial\Phi}{\partial\theta^k} = -s\langle H_k\rangle_\theta. \tag{11.15}$$

The Massieu function $\Phi(\theta)$, because it is a Legendre transform, is automatically a convex function of the parameters θ^k.

The thermodynamic entropy $S(U)$ is by definition the value attained by $S_\phi(p_\theta)$ when $\langle H_k\rangle_\theta = U_k$ for all k. A short calculation gives

$$\frac{\partial S}{\partial U_k} = s\theta^k. \tag{11.16}$$

See the Box 11.3. This relation is the *dual* of (11.15) — see (3.17, 3.18).

Finally, note that $S(U)$ is the inverse Legendre transform of $\phi(\theta)$. In formulae,

$$S(U) = \inf_\theta\{\Phi(\theta) + s\theta^k U_k\}. \tag{11.20}$$

This implies that $S(U)$ is a concave function of U, showing that the ϕ-exponential probability distributions are always thermodynamically stable.

First calculate, using (10.38),

$$\begin{aligned}\frac{\partial}{\partial\theta^k}S_\phi(p_\theta) &= -s\sum_j \ln_\phi(p_\theta(j))\frac{\partial}{\partial\theta^k}p_\theta(j)\\ &= -s\sum_j \left[-\alpha(\theta)-\theta^l H_l(j)\right] z(\theta)P_\theta(j)\left[-\frac{\partial\alpha}{\partial\theta^k}-H_k(j)\right]\\ &= -sz(\theta)\theta^l g_{lk}(\theta), \end{aligned}\tag{11.17}$$

with

$$g_{lk}(\theta) = \langle\langle H_k H_l\rangle\rangle_\theta - \langle\langle H_k\rangle\rangle_\theta\,\langle\langle H_l\rangle\rangle_\theta. \tag{11.18}$$

Next calculate, also using (10.38),

$$\begin{aligned}\frac{\partial U_l}{\partial\theta^k} &= \frac{\partial}{\partial\theta^k}\sum_j p_\theta(j)H_l(j)\\ &= z(\theta)\sum_j P_\theta(j)\left[-\frac{\partial\alpha(\theta)}{\partial\theta^k}-H_k(j)\right]H_l(j)\\ &= -z(\theta)g_{kl}. \end{aligned}\tag{11.19}$$

Combining both results yields (11.16).

Box 11.3 The derivatives of the thermodynamic entropy

11.4 Complexity

In the context of *game theory* the probability distributions p and q are the strategies of two players. The *relative complexity* $C(p||q)$ of these strategies has been defined [6, 15] as

$$C(p||q) = S(p) + D(p||q). \tag{11.21}$$

The basic property of a complexity $C(p||q)$ is that $C(p||q) \geq C(q||q)$, with equality if and only if $p = q$. From the complexity function one can deduce the entropy functional $S(p)$ by taking both arguments equal: $S(p) = C(p||p)$. Next, the relative entropy follows as the difference between actual complexity and minimal complexity. Hence, complexity is the more fundamental of the three quantities.

From the definitions of ϕ-deformed entropy and relative entropy follows (see (11.10)

$$\begin{aligned}C_\phi(p||q) &= S_\phi(p) + D_\phi(p||q)\\ &= S_\phi(q) - s\sum_j [p(j)-q(j)]\ln_\phi(q(j)). \end{aligned}\tag{11.22}$$

In particular, when q is a member of the ϕ-exponential family, this is, $q = p_\theta$ as given by (10.30) (assuming $c(x) = 1$), then one obtains

$$
\begin{aligned}
C_\phi(p||p_\theta) &= S_\phi(p_\theta) + s\sum_j [p(j) - p_\theta(j)]\theta^k H_k(j) \\
&= \Phi(\theta) + s\theta^k \langle H\rangle_p.
\end{aligned}
\tag{11.23}
$$

As discussed before in Section 3.1, the variational principle expresses the *stability* of the physical system with respect to perturbations. In the context of game theory the relevant notion is that of robustness. By definition, the strategy q is *robust* within the set $\mathcal{P}$ of strategies if $C(p||q)$ has a constant value not depending on the choice of p in $\mathcal{P}$. From (11.23) then follows that each member p_θ of the ϕ-exponential family is robust within the set of strategies p which have the same expectation $\langle H_k\rangle_p$ of the estimators H_k as p_θ.

The complexities $C_\phi(p||q)$ can be written as (see (11.22))

$$
C_\phi(p||q) = \xi(q) + \langle \kappa^{(q)}\rangle_p,
\tag{11.24}
$$

with

$$
\begin{aligned}
\xi(q) &= S_\phi(q) + s\sum_k q(k)\ln_\phi(q(k)) \\
\kappa^{(q)}(j) &= -s\ln_\phi(q(j)).
\end{aligned}
\tag{11.25}
$$

In the undeformed case ($\phi(u) = u$) is $S_\phi(q) = -\sum_k q(k)\ln q(k)$ so that in this case $\xi(q) = 0$. Then the only contribution is the expectation of $\kappa^{(q)}(j) = -\ln(q(j))$. This is the amount of information encoded in event j. The function $\kappa(u) = -\ln u$ is the *coding function*. Hence, $\langle\kappa^{(q)}\rangle_p$ is the coding contribution to the complexity. The term $\xi(q)$ has been called the *corrector*.

11.5 Application to q-Deformed Distributions

Let us now apply the theory to the case that $\phi(u) = u^q$ with $0 < q < 2$. The condition $q < 2$ is needed to assure the existence of the deduced logarithm $\omega_\phi(u)$. The condition $q > 0$ is needed to make that $\phi(u)$ is an increasing function. The corresponding deformed logarithmic and exponential functions are those introduced in the Chapter 7. The q-exponential family was studied in the Section 7.3, the escort family in the Section 7.4.

The constant s is given by

$$
\frac{1}{s} = \int_0^1 \mathrm{d}v\,\frac{v}{v^q} = \frac{1}{2-q}.
\tag{11.26}
$$

The deduced logarithm $\omega_\phi(u)$ equals

$$\begin{aligned}\omega_q(u) &= u\int_0^{1/u} \mathrm{d}v\, \frac{v}{v^q} - \frac{1}{s} - \ln_q \frac{1}{u}\\ &= \frac{1}{s}u^{q-1} - \frac{1}{s} - \ln_q \frac{1}{u}\\ &= -\frac{1}{s}\ln_q \frac{1}{u}. \end{aligned} \tag{11.27}$$

The entropy functional $S_\phi(p)$ becomes (see (8.3))

$$\begin{aligned}S_q(p) &= s\sum_j p(j)\omega_q\left(\frac{1}{p(j)}\right)\\ &= -\sum_j p(j)\ln_q p(j)\\ &= S^{\mathrm{Tsallis}}_{2-q}(p). \end{aligned} \tag{11.28}$$

The thermodynamic entropy $S(U)$ therefore equals

$$\begin{aligned}S(U) &= S_q(p_\theta)\\ &= -s\sum_j p(j)\left[-\alpha(\theta) - \theta^k H_k\right]\\ &= s\left[\alpha(\theta) + \theta^k U_k\right], \end{aligned} \tag{11.29}$$

with $U_k = \langle H_k\rangle_\theta$. This implies (8.9).

The relative entropy of the Bregman type $D_\phi(p||q)$ reduces to (8.49) (Problem 11.1).

11.6 Deformed Fisher Information

In the present Section the assumption $s = 1$ is made — see (11.2 for the definition of the constant s. Note that this is not a loss of generality since $s = 1$ can be arranged by multiplying the function $\phi(u)$ with a constant.

Note that the matrix $g_{kl}(\theta)$, introduced during the calculations found in the Box 11.3, is positive-definite (its eigenvalues cannot be negative). It therefore has the meaning of a *metric tensor*. associating with each choice of the parameters θ^k a point of a manifold. From (11.19) and (11.15) follows

$$\frac{\partial^2\Phi}{\partial\theta^k\partial\theta^l} = z(\theta)g_{kl}(\theta). \tag{11.30}$$

Hence, the geometry of the manifold, determined by the value of the Massieu function $\Phi(\theta)$, is characterised by the metric tensor $z(\theta)g_{kl}$. The relevance of this observation follows from the link with information theory and is worked out below.

In the present context of parametrised models of statistical physics the Hamiltonians H_k are used to estimate the parameters θ^k of the model. More precisely, it was pointed out in the previous Chapter that the *escort probabilities* $\langle\langle H_k \rangle\rangle_\theta$ can be used to this purpose. When they are known then the derivatives of the Massieu function $\Phi(\theta)$ are known via (11.15). By constructing a tangent plane with these derivatives one obtains the parameter θ form the knowledge of the tangent point. The accuracy by which this tangent point θ can be determined depends on the curvature of the manifold at θ, this is, on the metric tensor $g_{kl}(\theta)$. The more curved the surface is, the more accurate is the determination of the tangent point.

A mathematical expression of this relation is the *inequality of Cramér and Rao*. Its generalisation to pairs of families of probability distributions reads

Theorem 11.1. *Let $p_\theta(j)$ and $P_\theta(j)$ be two families of probability distributions with a common domain of definition. Let $\langle \cdot \rangle_\theta$ and $\langle\langle \cdot \rangle\rangle_\theta$ denote the averages with respect to $p_\theta(j)$, respectively $P_\theta(j)$. Let be given Hamiltonians H_k and a function $\Phi(\theta)$ such that*

$$\frac{\partial \Phi}{\partial \theta^k} = -\langle H_k \rangle_\theta. \tag{11.31}$$

Introduce the information matrix

$$I_{kl}(\theta) = \sum_j \frac{1}{P_\theta(j)} \frac{\partial p_\theta(j)}{\partial \theta^k} \frac{\partial p_\theta(j)}{\partial \theta^l}. \tag{11.32}$$

Then for all choices of the real numbers u^k and v^k is

$$\frac{u^k u^l \left[\langle\langle H_k H_l \rangle\rangle_\theta - \langle\langle H_k \rangle\rangle_\theta \langle\langle H_l \rangle\rangle_\theta\right]}{\left[u^k v^l \frac{\partial^2 \Phi}{\partial \theta^k \partial \theta^l}\right]^2} \geq \frac{1}{v^k v^l I_{kl}(\theta)}. \tag{11.33}$$

If $p_\theta(j)$ belongs to the ϕ-exponential family and $P_\theta(j)$ is the corresponding escort family then the inequality becomes an equality whenever $u = v$. In that case one has

$$I_{kl}(\theta) = z(\theta)^2 g_{kl}(\theta). \tag{11.34}$$

The proof is given in the Box 11.4.

The expression (11.32) can be seen as a generalisation of the *Fisher information* matrix. The quantity $\langle\langle H_k H_l \rangle\rangle_\theta - \langle\langle H_k \rangle\rangle_\theta \langle\langle H_l \rangle\rangle_\theta$ is usually interpreted as a measure for the inaccuracy of the estimated values $\langle\langle H_k \rangle\rangle_\theta$. In the case of unbiased estimators the denominator of the l.h.s. reduces to $(u^k v_k)^2$. Then the inequality gives a lower bound for the inaccuracy of the estimation. The knowledge of the r.h.s. is then important to have an idea whether the es-

Let $X_k(j) = \frac{1}{P_\theta(j)}\frac{\partial p_\theta(j)}{\partial\theta^k}$ and $Y_k(j) = H_k(j) - \langle\langle H_k\rangle\rangle_\theta$. Then Schwartz's inequality implies $\left(\langle\langle u^k Y_k v^l X_l\rangle\rangle_\theta\right)^2 \le \langle\langle u^k Y_k u^l Y_l\rangle\rangle_\theta \langle\langle v^k X_k v^l X_l\rangle\rangle_\theta$. The l.h.s. can be written as

$$\begin{aligned}\left(\langle\langle u^k Y_k v^l X_l\rangle\rangle_\theta\right)^2 &= \left(u^k v^l \sum_j [H_k(j) - \langle\langle H_k\rangle\rangle_\theta]\frac{\partial p_\theta(j)}{\partial\theta^l}\right)^2\\ &= \left(u^k v^l \frac{\partial}{\partial\theta^l}\langle H_k\rangle_\theta\right)^2 = \left(u^k v^l \frac{\partial^2 \Phi}{\partial\theta^k\partial\theta^l}\right)^2 . \qquad (11.35)\end{aligned}$$

The first factor of the r.h.s. equals

$$u^k u^l \left[\langle\langle H_k H_l\rangle\rangle_\theta - \langle\langle H_k\rangle\rangle_\theta \, \langle\langle H_l\rangle\rangle_\theta\right]. \qquad (11.36)$$

The second factor equals

$$v^k v^l \langle\langle X_k X_l\rangle\rangle_\theta = v^k v^l \sum_j \frac{1}{P_\theta(j)}\frac{\partial p_\theta(j)}{\partial\theta^k}\frac{\partial p_\theta(j)}{\partial\theta^l} = v^k v^l I_{kl}(\theta). \qquad (11.37)$$

Hence, the three pieces together prove the inequality (11.33).

Assume now that $p_\theta(j)$ belongs to the ϕ-exponential family and $P_\theta(j)$ is the corresponding escort family. Then one calculates using (10.38)

$$\begin{aligned}I_{kl}(\theta) &= z(\theta)^2 \sum_j P_\theta(j)\left(-\frac{\partial\alpha}{\partial\theta^k} - H_k(j)\right)\left(-\frac{\partial\alpha}{\partial\theta^l} - H_l(j)\right)\\ &= z(\theta)^2 g_{kl}(\theta). \qquad (11.38)\end{aligned}$$

This proves (11.34). Using in addition (11.30), the inequality(11.33) becomes

$$\frac{u^k u^l g_k l(\theta)}{z(\theta)^2\left[u^k v^l g_{kl}(\theta)\right]^2} \ge \frac{1}{z(\theta)^2 v^k v^l g_{kl}(\theta)}, \qquad (11.39)$$

which indeed becomes an equality when $u = v$.

Box 11.4 Proof of the generalised inequality of Cramér and Rao

timation is optimal. In the case of a ϕ-exponential distribution, the accuracy is optimal when the parameters θ^k are estimated using the escort family.

11.7 Quantum Entropies

Fix a non-decreasing function $\phi(u)$, defined for $u \ge 0$, and strictly positive for $u > 0$. Let $\omega_\phi(u)$ be the deduced logarithm. Then the classical definition (11.1) suggests to define the quantum ϕ-deformed entropy functional by

$$S_\phi(\rho) = s\,\mathrm{Tr}\,\rho\omega_\phi(\rho^{-1}) = -s\,\mathrm{Tr}\,\rho\omega_\phi^*(\rho). \tag{11.40}$$

Remember that the constant s is defined by (11.2). The deduced logarithm $\omega_\phi(u)$ is positive for $u \geq 1$ while the eigenvalues of ρ are smaller than one. Hence, $\omega_\phi(\rho^{-1})$ is a positive operator. There follows that $0 \leq S_\phi(\rho) \leq +\infty$. The proof of the concavity of $S_\phi(\rho)$ is postponed to the end of this Section.

Expression (11.40) can be written as

$$S_\phi(\rho) = -s\int_0^1 \mathrm{d}u\ \mathrm{Tr}\,\rho\ln_\phi(u\rho) - 1. \tag{11.41}$$

To see this, note that

$$\begin{aligned}
\omega_\phi\left(\frac{1}{u}\right) &= \frac{1}{u}\int_0^u \mathrm{d}v\,\frac{v}{\phi(v)} - \frac{1}{s} - \ln_\phi(u)\\
&= u\int_0^1 \mathrm{d}w\,\frac{w}{\phi(uw)} - \frac{1}{s} - \ln_\phi(u)\\
&= -\int_0^1 \mathrm{d}w\ \ln_\phi(uw) - \frac{1}{s}.
\end{aligned} \tag{11.42}$$

Partial integration was used to obtain the last line.

The corresponding quantum expression of the relative entropy is

$$D_\phi(\rho||\sigma) = S_\phi(\sigma) - S_\phi(\rho) - \ \mathrm{Tr}\,(\rho-\sigma)\ln_\phi(\sigma). \tag{11.43}$$

Its basic property is that $D_\phi(\rho||\sigma) \geq 0$. The difficulty in proving this inequality is that ρ and σ do not necessarily commute. Therefore, there need not exist a basis in which both are simultaneously diagonal. See the Box 11.5. An immediate consequence is that $S_\phi(\rho)$ is concave — see the Box 11.6.

11.8 Quantum Stability

Let ρ_θ be the quantum ϕ-exponential family, given by (10.62). Then one finds

$$D(\sigma||\rho_\theta) = S_\phi(\sigma) - S_\phi(\rho_\theta) + s\theta^k\,\mathrm{Tr}\,(\rho_\theta - \sigma)H_k. \tag{11.49}$$

The positivity of the relative entropy $D(\sigma||\rho_\theta) \geq 0$ then implies the *variational principle*, stating that the non-equilibrium Massieu functional

$$S_\phi(\sigma) - s\theta^k\,\mathrm{Tr}\,\sigma H_k \tag{11.50}$$

is minimal when $\sigma = \rho_\theta$.

The Massieu function is defined by

$$\Phi(\theta) = S_\phi(\rho_\theta) - \theta^k\,\mathrm{Tr}\,\rho_\theta H_k. \tag{11.51}$$

The following result is known as Klein's inequality.

Lemma Let A, B and C be self-adjoint operators with discrete spectrum. Assume that $C \geq 0$ and $BC = CB$. Let $f(x)$ be a convex function. Then one has

$$\operatorname{Tr} C[f(A) - f(B) - (A - B)f'(B)] \geq 0. \tag{11.44}$$

Proof Let $(\phi_n)_n$ be an orthonormal basis in which A is diagonal. Let $(\psi_m)_m$ be an orthonormal basis in which B and C are simultaneously diagonal. Let $A\phi_n = a_n\phi_n$, $B\psi_m = b_m\psi_m$, and $C\psi_m = c_m\psi_m$. Denote $\lambda_{nm} = \langle\phi_m|\psi_n\rangle$. Then the convexity of $f(x)$ implies that

$$\begin{aligned} &\langle\phi_m|C(f(A) - f(B) - (A - B)f'(B))|\phi_m\rangle \\ &= \sum_n c_n|\lambda_{mn}|^2[f(a_m) - f(b_n) - (a_m - b_n)f'(b_n)] \\ &\geq 0. \end{aligned} \tag{11.45}$$

To see the inequality, use that a tangent line to a convex function always lies below the function.

Klein's inequality now follows by summing over m.

Consider now the function $f(u) = -u\omega_\phi\left(\frac{1}{u}\right)$, in combination with $C = \mathbf{I}$, $A = \rho$, and $B = \sigma$. From the basic property of deduced logarithms, (10.25), follows $f'(u) = \ln_\phi(u) + F_\phi(0)$, so that $f''(u) = 1/\phi(u) > 0$. Hence, $f(u)$ is convex and Klein's inequality may be applied. The result is

$$-S_\phi(\rho) + S_\phi(\sigma) - s\operatorname{Tr}(\rho - \sigma)\ln_\phi(\sigma) \geq 0. \tag{11.46}$$

Box 11.5 Positivity of the relative entropy

Let $\tau = \lambda\sigma + (1 - \lambda)\rho$ with $0 \leq \lambda \leq 1$. One has

$$\begin{aligned} 0 \leq D(\rho||\tau) &= S_\phi(\tau) - S_\phi(\rho) - s\operatorname{Tr}(\rho - \tau)\ln_\phi(\tau), \\ 0 \leq D(\sigma||\tau) &= S_\phi(\tau) - S_\phi(\sigma) - s\operatorname{Tr}(\sigma - \tau)\ln_\phi(\tau). \end{aligned} \tag{11.47}$$

By multiplying the first equation with $1 - \lambda$, the second with λ, and summing, one obtains

$$0 \leq S_\phi(\tau) - \lambda S_\phi(\sigma) - (1 - \lambda)S_\phi(\rho). \tag{11.48}$$

This proves the concavity of the entropy functional $S_\phi(\rho)$.

Box 11.6 Concavity of the quantum entropy functional

Using (10.25) and the invariance of the trace $\mathrm{Tr}\,AB = \mathrm{Tr}\,BA$, one obtains

$$\begin{aligned}\frac{\partial}{\partial\theta^k}S_\phi(\rho_\theta) &= s\frac{\partial}{\partial\theta^k}\,\mathrm{Tr}\,\rho_\theta\omega_\phi(\rho_\theta^{-1})\\ &= \mathrm{Tr}\,[-s\ln_\phi(\rho_\theta)-1]\,\frac{\partial}{\partial\theta^k}\rho_\theta\\ &= \mathrm{Tr}\,\left[s\alpha(\theta)+s\theta^l H_l-1\right]\frac{\partial}{\partial\theta^k}\rho_\theta\\ &= s\theta^l\frac{\partial}{\partial\theta^k}\,\mathrm{Tr}\,\rho_\theta H_l.\end{aligned} \tag{11.52}$$

Therefore one has, as expected,

$$\frac{\partial\Phi}{\partial\theta^k} = -s\,\mathrm{Tr}\,\rho_\theta H_k \equiv -s\langle H_k\rangle_\theta. \tag{11.53}$$

Problems

11.1. q-deformed relative entropy
Show that the q-deformed relative entropy (8.49), introduced in Chapter 8 is a special case of the ϕ-deformed relative entropy (11.9).

11.2. A two-parameter family
Borges and Roditi [4] introduced into the physics literature the entropy functional

$$S_{q,r}(p) = \sum_j \frac{p(j)^r - p(j)^q}{q-r} \tag{11.54}$$

(see the Section 8.9). It is a ϕ-deformed entropy functional with $\phi(u)$ given by

$$\phi(u) = \frac{q-r}{q(q-1)u^{q-2} - r(r-1)u^{r-2}}. \tag{11.55}$$

The latter is a positive increasing function of $u \in (0,+\infty)$ when $0 \le r \le 1 \le q \le 2$ or $0 \le q \le 1 \le r \le 2$. The corresponding deformed logarithm is

$$\ln_\phi(u) = \frac{qu^{q-1} - ru^{r-1}}{q-r} - 1. \tag{11.56}$$

Verify these statements.

11.3. Kaniadakis' entropy functional
The deformed exponential and logarithmic functions of Kaniadakis have been introduced in the Box 10.1. On the other hand, the entropy function postulated by Giorgio Kaniadakis and Antonio Scarfone [7] is a special case of the

two-parameter family discussed in the previous Problem. It reads

$$S_\kappa(f) = \frac{1}{2\kappa}\int \mathrm{d}x\, \left(f(x)^{1-\kappa} - f(x)^{1+\kappa}\right) = \int \mathrm{d}x\, f(x) \ln_\kappa\left(\frac{1}{f(x)}\right) \quad (11.57)$$

This means, in the formalism of the present Chapter, that $\ln_\kappa(u)$ is the *deduced logarithm* $\omega_\phi(u)$ corresponding with some monotone function $\phi(u)$. Determine this function and show that it satisfies the relation $\phi'(u) = \frac{1-\kappa}{u}\phi(u)$.

Notes

Most of this chapter is based on my papers [8, 9]. I learned about Bregman type relative entropy from Christophe Vignat and proposed its generalisation in [10]. The mathematical properties of these generalised entropy functionals are discussed in more detail in these papers.

The definition of entropy functional used by Chavanis [5] is slightly different from the one introduced here. In (11.4), he replaces the integrand $\ln_\phi(y)$ by an arbitrary increasing function, not necessarily concave.

The discussion of Fisher' s information in the context of q-deformed entropies goes back to Sumiyoshi Abe [1, 2] and Pennini et al [12]. The geometric aspects of the exponential family are known in the mathematics literature under the name of geometry of the *statistical manifold* [3]. The connection with Amari's information geometry was the starting point of [8] and is worked out in [11].

Klein's inequality can be found for instance in [13], 2.5.2, or [14], 2.1.7.

Objectives

- Know the definition of the generalised entropy functional in terms of the deduced logarithm.
- Use the relative entropy of the Bregman type to prove the thermodynamic stability of the phi-exponential family.
- Apply the general formalism to the case that $\phi(u) = u^q$. Know why q is limited to lie between 0 and 2.
- Discuss the Fisher information and the inequality of Cramér and Rao in the context of phi-deformed functions.
- Translate the classical expressions into quantum mechanical expressions.
- Know about Klein's inequality.

References

1. Abe, S.: q-deformed entropies and Fisher metrics. In: P. Kasperkovitz, D. Grau (eds.) Proceedings of The 5th International Wigner Symposium, August 25–29, 1997, Vienna, Austria, p. 66. World Scientific, Singapore (1998) 123, 177
2. Abe, S.: Geometry of escort distributions. Phys. Rev. E **68**, 031101 (2003) 177
3. Amari, S.: Differential-geometrical methods in statistics, *Lecture Notes in Statistics*, vol. 28. Springer, New York, Berlin (1985) 177
4. Borges, E., Roditi, I.: A family of nonextensive entropies. Phys. Lett. A **246**, 399–402 (1998) 128, 176
5. Chavanis, P.H.: Generalized thermodynamics and Fokker-Planck equations: Applications to stellar dynamics and two-dimensional turbulence. Phys. Rev. E **68**, 036108 (2003) 177
6. Harremoës, P., Topsøe, F.: Maximum entropy fundamentals. Entropy **3**, 191–226 (2001) 169
7. Kaniadakis, G., Scarfone, A.: A new one parameter deformation of the exponential function. Physica A **305**, 69–75 (2002) 128, 152, 176
8. Naudts, J.: Estimators, escort probabilities, and phi-exponential families in statistical physics. J. Ineq. Pure Appl. Math. **5**, 102 (2004) 112, 162, 177
9. Naudts, J.: Escort operators and generalized quantum information measures. Open Systems and Information Dynamics **12**, 13–22 (2005) 162, 177
10. Naudts, J.: Continuity of a class of entropies and relative entropies. Rev. Math. Phys. **16**, 809–822 (2004); Errata. Rev. Math. Phys. **21**, 947–948 (2009) 129, 162, 177
11. Ohara, A., Matsuzoe, H., Amari, S.: A dually flat structure on the space of escort distributions. J. Phys.: Conf. Ser. **201**, 012012 (2010) 177
12. Pennini, F., Plastino, A., Plastino, A.: Rényi entropies and Fisher informations as measures of nonextensivity in a Tsallis setting. Physica **258**, 446–457 (1998) 177
13. Ruelle, D.: Statistical mechanics, Rigorous results. W.A. Benjamin, Inc., New York (1969) 177
14. Thirring, W.: Lehrbuch der Mathematischen Physik, Vol. 4. Springer-Verlag, Wien (1980) 177
15. Topsøe, F.: Entropy and equilibrium via games of complexity. Physica A **340**, 11–31 (2004) 169

Solutions to the Problems

Problems of Chapter 1

1.1 Experimental determination of Boltzmann's constant
Let z_j denote the height of the j-th particle above the bottom of the cylindrical vessel. The potential energy of N particles is then $\sum_{j=1}^N mgz_j$. The probability density of the altitudes of the particles is then

$$p(z_1, z_2, \cdots, z_N) = \frac{1}{Z(\beta)} e^{-\beta \sum_{j=1}^N mgz_j}. \tag{12.1}$$

The normalisation $Z(\beta)$ is given by

$$\begin{aligned} Z(\beta) &= \prod_{j=1}^N \left(\int_0^\infty \mathrm{d}z_j \right) \exp(-\beta \sum_{j=1}^N mgz_j) \\ &= \left(\int_0^{+\infty} \mathrm{d}z \exp(-\beta mgz) \right)^N \\ &= (\beta mg)^{-N}. \end{aligned} \tag{12.2}$$

The expected number of particles between altitudes h and $h + \Delta h$ ($\Delta h > 0$) is then

$$\begin{aligned} &\prod_{j=1}^N \left(\int_0^\infty \mathrm{d}z_j \right) \sum_{j=1}^N \mathbf{I}_{(h \le z_j \le h+\Delta h)} p(z_1, \cdots z_N) \\ &= N\beta mg \int_0^\infty \mathrm{d}z \, \mathbf{I}_{(h \le z_j \le h+\Delta h)} \exp(-\beta mgz) \\ &= N \left[\exp(-\beta mgh) - \exp(-\beta mg(h + \Delta h)) \right] \\ &= N \exp(-\beta mgh) \left[1 - \exp(-\beta mg\Delta h) \right]. \end{aligned} \tag{12.3}$$

The logarithm of this expression is of the form constant $- \beta mgh$. By fitting to the logarithm of the measured particle counts one gets an estimate for

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8_12,

the slope βmg. See the Figure 2.1. Since m and g are known this yields $\beta = 1/k_{\rm B}T$. But also T is known. Hence one obtains an estimate for k_B. Using a spreadsheet one can do the fit quickly. The result of 1.4×10^{-23} J/K is not far from the best known value, which is about 1.38065×10^{-23} J/K.

1.2 A quantum spin

Introduce the notation $\phi_t = \mu t/\hbar$. One has

$$U(t) = e^{-\frac{i}{\hbar}tH} = e^{\frac{i}{\hbar}\mu t\sigma_z} = \cos(\phi_t)\mathbf{I} + i\sin(\phi_t)\sigma_z. \tag{12.4}$$

The latter step follows because $\sigma_z^2 = \mathbf{I}$. Therefore

$$\begin{aligned}\sigma_x(t) &= U(t)^\dagger \sigma_x U(t)\\ &= U(t)^\dagger \sigma_x \left[\cos(\phi_t)\mathbf{I} + i\sin(\phi_t)\sigma_z\right]\\ &= \left[\cos(\phi_t)\mathbf{I} - i\sin(\phi_t)\sigma_z.\right]\left[\cos(\phi_t)\sigma_x + \sin(\phi_t)\sigma_y\right]\\ &= \cos(2\phi_t)\sigma_x + \sin(2\phi_t)\sigma_y.\end{aligned} \tag{12.5}$$

On the other hand is

$$\rho = \frac{e^{-\beta H}}{\mathrm{Tr}\, e^{-\beta H}} = \frac{e^{\beta\mu]\sigma_z}}{\mathrm{Tr}\, e^{\beta\mu]\sigma_z}} = \frac{\cosh(\beta\mu)\mathbf{I} + \sinh(\beta\mu)\sigma_z}{2\cosh(\beta\mu)}. \tag{12.6}$$

All together is

$$\begin{aligned}\langle\sigma_x(t) = \mathrm{Tr}\,\rho\sigma_x(t) &= \mathrm{Tr}\left[\frac{\cosh(\beta\mu)\mathbf{I} + \sinh(\beta\mu)\sigma_z}{2\cosh(\beta\mu)}\right]\left[\cos(2\phi_t)\sigma_x + \sin(2\phi_t)\sigma_y\right]\\ &= \mathrm{Tr}\left[\cos(2\phi_t)\sigma_x + \sin(2\phi_t)\sigma_y\right]\\ &\quad + i\tanh(\beta\mu)\,\mathrm{Tr}\left[\cos(2\phi_t)\sigma_y - \sin(2\phi_t)\sigma_x\right]\\ &= 0.\end{aligned} \tag{12.7}$$

1.3 A quantum particle trapped between two walls

The partition sum equals

$$\begin{aligned}Z(\beta) = \sum_{n=0}^{\infty} e^{-\beta E_n} &= \sum_{n=0}^{\infty} \exp\left(-\beta\frac{\hbar^2\pi^2}{2mL^2}n^2\right)\\ &\simeq \int_0^\infty \mathrm{d}x\, \exp\left(-\beta\frac{\hbar^2\pi^2}{2mL^2}x^2\right)\\ &= \frac{L\sqrt{m}}{\hbar\sqrt{2\pi\beta}}.\end{aligned} \tag{12.8}$$

The average energy then follows from

$$\begin{aligned}U(T) = \langle H\rangle &= -\frac{\mathrm{d}}{\mathrm{d}\beta}\ln Z(\beta)\\ &= \frac{1}{2\beta} = \frac{1}{2}k_{\rm B}T.\end{aligned} \tag{12.9}$$

1.4 Two fermions in the canonical ensemble
One of the fermions is in the eigenstate m, the other in the eigenstate n, with $n \neq m$. Hence one has

$$Z_2(\beta) = \sum_{m=0}^{\infty} \sum_{n=m+1}^{\infty} e^{-\beta(\epsilon_m+\epsilon_n)} = e^{-\beta\hbar\omega} \sum_{m=0}^{\infty} \sum_{n=m+1}^{\infty} e^{-\beta\hbar\omega(m+n)}. \quad (12.10)$$

Introduce the new summing variable $p = n - m - 1$. Then one gets

$$\begin{aligned} Z_2(\beta) &= e^{-\beta\hbar\omega} \sum_{m,p=0}^{\infty} e^{-\beta\hbar\omega(m+p+m+1)} \\ &= e^{-2\beta\hbar\omega} \sum_{m=0}^{\infty} e^{-2\beta\hbar\omega m} \sum_{p=0}^{\infty} e^{-\beta\hbar\omega p} \\ &= e^{-2\beta\hbar\omega} \frac{1}{1-\exp(-2\beta\hbar\omega)} \frac{1}{1-\exp(-\beta\hbar\omega)} \\ &= \frac{1}{1+\exp(-\beta\hbar\omega)} \left(\frac{1}{\exp(\beta\hbar\omega)-1}\right)^2 . \end{aligned} \quad (12.11)$$

Problems of Chapter 2

2.1 Correlations in the one-dimensional Ising model
Introduce new variables $\tau_n = \sigma_n\sigma_{n+1}$, $n = 1, 2, \cdots N$. Then the partition sum $Z_N(\beta, 0)$ can be evaluated as follows

$$\begin{aligned} Z_N(\beta,0) &= \sum_{\sigma_1=\pm1} \cdots \sum_{\sigma_N=\pm1} \exp\left(\beta J \sum_{n=1}^{N} \sigma_n\sigma_{n+1}\right) \\ &= \sum_{\tau_1=\pm1} \cdots \sum_{\tau_N=\pm1} \exp\left(\beta J \sum_{n=1}^{N} \tau_n\right) \\ &= \prod_{n=1}^{N} \sum_{\tau_n=\pm1} \exp\left(\beta J \tau_n\right) \\ &= 2^N \left(\cosh(\beta J)\right)^N . \end{aligned} \quad (12.12)$$

Next note that one can write for $n > 1$

$$\sigma_1\sigma_n = \prod_{m=1}^{n-1} \tau_m, \quad (12.13)$$

so that

$$
\begin{aligned}
\langle \sigma_1 \sigma_n \rangle &= \frac{1}{Z_N(\beta,0)} \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_N=\pm 1} \exp\left(\beta J \sum_{p=1}^{N} \sigma_p \sigma_{p+1}\right) \prod_{m=1}^{n-1} \tau_m \\
&= \frac{1}{Z_N(\beta,0)} \sum_{\tau_1=\pm 1} \cdots \sum_{\tau_N=\pm 1} \exp\left(\beta J \sum_{p=1}^{N} \tau_p\right) \prod_{m=1}^{n-1} \tau_m \\
&= \frac{\prod_{m=1}^{n-1} \sum_{\tau_m=\pm 1} \tau_m \exp\left(\beta J \tau_m\right)}{\prod_{m=1}^{n-1} \sum_{\tau_m=\pm 1} \exp\left(\beta J \tau_m\right)} \\
&= \left(\tanh(\beta J)\right)^{n-1} .
\end{aligned}
\tag{12.14}
$$

2.2 The Gamma distribution

It is obvious to write

$$
p(x) = \frac{x^{k-1} e^{-x/b}}{b^k \Gamma(k)} = \exp\left(-k \ln b - \ln \Gamma(k) + (k-1)\ln x - \frac{x}{b}\right). \tag{12.15}
$$

Introduce the Hamiltonians $H_1(x) = \ln x$ and $H_2(x) = x$. Then, using the proposed definition of the parameters θ_1 and θ_2, the density $p(x)$ has the right form for a member of the exponential family. The Massieu function $\Phi(\theta)$ equals

$$
\Phi(\theta) = k \ln b + \ln \Gamma(k) = (\theta_1 - 1)\ln \theta_2 + \ln \Gamma(1 - \theta_1). \tag{12.16}
$$

2.3 Example of the quantum exponential family

Let us try to write

$$
\rho_\lambda = \exp(-\Phi(\theta) - \theta H) \quad \text{with} \quad H = -\sigma_z. \tag{12.17}
$$

From the identification of

$$
\rho = \frac{1}{2} + \left(\lambda - \frac{1}{2}\right)\sigma_3 \tag{12.18}
$$

with

$$
\exp(-\Phi(\theta) - \theta H) = e^{-\Phi(\theta)}\left(\cosh(\theta) + \sigma_z \sinh(\theta)\right) \tag{12.19}
$$

follows that $\tanh(\theta) = 2\lambda - 1$ and $\Phi(\theta) = \ln(2\cosh(\theta))$. This shows that it is possible to write ρ_λ in the form (12.17).

2.4 Density profile of the earth

The Massieu function is given by

$$
\begin{aligned}
\Phi(\theta) &= \ln \int_0^\infty \mathrm{d}\rho_1 \cdots \int_0^\infty \mathrm{d}\rho_N \, \exp\left(-\theta_1 H_1(\rho) - \theta_2 H_2(\rho)\right) \\
&= \sum_{n=1}^{N} \ln \int_0^\infty \mathrm{d}\rho_n \, \exp\left(-\eta_1 \rho_n - \eta_2 (n^{5/3} - (n-1)^{5/3})\rho_n\right)
\end{aligned}
$$

$$= -\sum_{n=1}^{N} \ln\left(\eta_1 + \eta_2(n^{5/3} - (n-1)^{5/3})\right). \tag{12.20}$$

with $\eta_1 = \theta_1/N$ and $\eta_2 = \theta_2/N^{5/3}$. There follows

$$\begin{aligned}\langle H_1\rangle_\theta &= -\frac{\partial\phi}{\partial\theta^1} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\eta_1 + \eta_2(n^{5/3} - (n-1)^{5/3})}\\ \langle H_2\rangle_\theta &= -\frac{\partial\phi}{\partial\theta^2} = \frac{1}{N^{5/3}}\sum_{n=1}^{N}\frac{n^{5/3} - (n-1)^{5/3}}{\eta_1 + \eta_2(n^{5/3} - (n-1)^{5/3})}.\end{aligned} \tag{12.21}$$

In the limit of large N this becomes

$$\begin{aligned}\langle H_1\rangle_\theta &= N\int_0^1 \mathrm{d}u\,\frac{1}{\theta_1 + \frac{5}{3}\theta_2 u^{2/3}}\\ \langle H_2\rangle_\theta &= \frac{5N}{3}\int_0^1 \mathrm{d}u\,\frac{u^{2/3}}{\theta_1 + \frac{5}{3}\theta_2 u^{2/3}}\end{aligned} \tag{12.22}$$

Note that $\theta_1\langle H_1\rangle_\theta + \theta_2\langle H_2\rangle_\theta = N$. Using

$$\int_0^1 \mathrm{d}u\,\frac{1}{a + bu^{2/3}} = \frac{3}{b}\left[1 - \sqrt{\frac{a}{b}}\arctan\sqrt{\frac{b}{a}}\right] \tag{12.23}$$

one finds

$$\langle H_1\rangle_\theta = \frac{9N}{5\theta_2}\left[1 - \sqrt{\frac{3\theta_1}{5\theta_2}}\arctan\sqrt{\frac{5\theta_2}{3\theta_1}}\right]. \tag{12.24}$$

The numerical solution of these two equations gives $\theta_1/N \simeq 0.000058$ and $\theta_2/N \simeq 0.000147$.

The probability distribution for the inner shell is proportional to

$$\exp\left(-\left[\frac{\theta_1}{N} + \frac{\theta_2}{N^{5/3}}\right]\rho_1\right). \tag{12.25}$$

In the limit of large N the second contribution may be neglected. Then an exponential distribution results, with average value $N/\theta_1 \simeq 17140$ kg/m^3.

2.5 Binomial distribution

Write

$$p_a(m) = \binom{n}{m} e^{m\ln a + ((n-m)\ln(1-a)}. \tag{12.26}$$

This shows that $p_a(m)$ is of the form (2.12), with $c(m) = \begin{pmatrix} n \\ m \end{pmatrix}$, $H(m) = m$, $\theta = -\ln \frac{a}{1-a}$, and $\Phi(\theta) = -n \ln(1 + e^{-\theta})$.

2.6 Weibull distribution
Take $c(x) = x^{k-1}$. Then one writes

$$f(x) = c(x) \exp\left(\ln k - k \ln \lambda - \lambda^{-k} x^k\right). \tag{12.27}$$

It is now obvious to choose $H(x) = x^k$ and $\theta = \lambda^{-k}$. Then the distribution takes on the form needed for a member of the exponential family.

Problems of Chapter 3

3.1 Binomial distribution revisited
The quantity to be optimised is (choose units for which $k_{\mathrm{B}} = 1$)

$$\mathcal{L}(p) = -\sum_{n=0}^{N} p(n) \ln \frac{p(n)}{c(n)} - \alpha \sum_{n=0}^{N} p(n) - \beta \sum_{n=0}^{N} p(n) n. \tag{12.28}$$

Variation w.r.t. $p(n)$ yields

$$0 = -\ln \frac{p(n)}{c(n)} - 1 - \alpha - \beta n. \tag{12.29}$$

This can be written as

$$p(n) = c(n) e^{-1-\alpha-\beta n}. \tag{12.30}$$

The parameters α and β must be fixed in such a way that the conditions $\sum_{n=0}^{N} p(n) = 1$ and $\sum_{n=0}^{N} p(n) n = \overline{n}$ are satisfied.

Assume now that $c(n) = \begin{pmatrix} N \\ n \end{pmatrix}$. Then one evaluates

$$1 = \sum_{n=0}^{N} p(n) = \sum_{n=0}^{N} \begin{pmatrix} N \\ n \end{pmatrix} e^{-1-\alpha-\beta n} = e^{-1-\alpha} \left(1 + e^{-\beta}\right)^{N}. \tag{12.31}$$

Using this result one can write the probabilities $p(n)$ in the following way

$$p(n) = \begin{pmatrix} N \\ n \end{pmatrix} \frac{e^{-\beta n}}{\left(1 + e^{-\beta}\right)^{N}} = \begin{pmatrix} N \\ n \end{pmatrix} a^n (1-a)^{N-n}. \tag{12.32}$$

In the latter line the notation $a = \dfrac{e^{-\beta}}{1+e^{-\beta}}$ has been introduced. The result is indeed the binomial distribution.

3.2 ***q*-deformed distribution**
The quantity to be optimised is

$$\mathcal{L}(p) = \frac{1}{1-q}\left(\sum_n p(n)\left(\frac{p(n)}{c(n)}\right)^{q-1} - 1\right) - \alpha\sum_{n=0}^{N} p(n) - \beta\sum_{n=0}^{N} p(n)n. \quad (12.33)$$

Variation w.r.t. $p(n)$ yields

$$0 = \frac{q}{1-q}\left(\frac{p(n)}{c(n)}\right)^{q-1} - \alpha - \beta n. \quad (12.34)$$

This can be written as

$$p(n) = c(n)\left[\left(\frac{1}{q} - 1\right)(\alpha + \beta n)\right]^{\frac{1}{q-1}}. \quad (12.35)$$

The parameters α and β must be fixed in such a way that the conditions $\sum_{n=0}^{N} p(n) = 1$ and $\sum_{n=0}^{N} p(n)n = \overline{n}$ are satisfied. However, this now seems to be a difficult task, even when we make the choice $c(n) = \binom{N}{n}$.

3.3 Entropy in the Bloch representation
The evaluation of the operator function $-\rho_{\mathbf{r}} \ln \rho_{\mathbf{r}}$ is best done in a basis in which the density matrix $\rho_{\mathbf{r}}$ is diagonal. Using the explicit expression

$$\rho_{\mathbf{r}} = \frac{1}{2}\begin{pmatrix} 1 - r_3 & -r_1 + ir_2 \\ -r_1 - ir_2 & 1 + r_3 \end{pmatrix}, \quad (12.36)$$

it is straightforward to determine the eigenvalues of $\rho_{\mathbf{r}}$. The secular equation (this is, the characteristic equation) is

$$(1 - r_3 - \lambda)(1 + r_3 - \lambda) - (r_1 + ir_2)(r_1 - ir_2) = 0. \quad (12.37)$$

Hence, the eigenvalues are $\lambda = \frac{1}{2}(1 \pm |\mathbf{r}|)$. The entropy therefore equals

$$S(\rho_{\mathbf{r}}) = -\frac{1}{2}(1 + |\mathbf{r}|)\ln\frac{1}{2}(1 + |\mathbf{r}|) - \frac{1}{2}(1 - |\mathbf{r}|)\ln\frac{1}{2}(1 - |\mathbf{r}|). \quad (12.38)$$

Note that this entropy vanishes when $\mathbf{r}$ approaches the surface of the Bloch sphere and that its maximal value $\ln 2$ is reached when $\mathbf{r} = 0$.

3.4 Approximate product measure
Because q is a product measure one can write $q(\sigma_1, \sigma_2) = q_1(\sigma_1)q_2(\sigma_2)$. Then the relative entropy $D(p||q)$ becomes

$$
\begin{aligned}
D(p||q) &= \sum_{\sigma_1,\sigma_2=\pm 1} p(\sigma_1,\sigma_2)\ln\frac{p(\sigma_1,\sigma_2)}{q(\sigma_1,\sigma_2)} \\
&= \sum_{\sigma_1,\sigma_2=\pm 1} p(\sigma_1,\sigma_2)\ln\frac{p(\sigma_1,\sigma_2)}{q_1(\sigma_1)q_2(\sigma_2)} \\
&= p(+,+)\ln\frac{p(+,+)}{q_1(+)q_2(+)} + p(+,-)\ln\frac{p(+,-)}{q_1(+)(1-q_2(+))} \\
&\quad p(-,+)\ln\frac{p(-,+)}{(1-q_1(+))q_2(+)} + p(-,-)\ln\frac{p(-,-)}{(1-q_1(+))(1-q_2(+))}.
\end{aligned}
\tag{12.39}
$$

Variation w.r.t. $q_1(+)$ respectively $q_2(+)$ yields the pair of equations

$$0 = -\frac{p(+,+)+p(+,-)}{q_1(+)} + \frac{p(-,+)+p(-,-)}{1-q_1(+)}, \tag{12.40}$$

$$0 = -\frac{p(+,+)+p(-,+)}{q_2(+)} + \frac{p(+,-)+p(-,-)}{1-q_2(+)}. \tag{12.41}$$

The solution of this set of equations is

$$q_1(+) = \sqrt{p(+,+)+p(+,-)}, \tag{12.42}$$

$$q_2(+) = \sqrt{p(+,+)+p(-,+)}. \tag{12.43}$$

3.5 Maxwell relations

Because $G(T,p)$ is a contact transform of $U - TS + pV$, one has

$$\frac{\partial G}{\partial T} = -S \quad \text{and} \quad \frac{\partial G}{\partial p} = V. \tag{12.44}$$

From

$$\frac{\partial^2 G}{\partial p \partial T} = \frac{\partial^2 G}{\partial T \partial p} \tag{12.45}$$

now follows that

$$-\frac{\partial S}{\partial p} = \frac{\partial V}{\partial T}. \tag{12.46}$$

This is the desired result. Hence, a Maxwell relation is nothing but a consequence of Clairaut's theorem (or Schwarz's theorem) that the order of partial derivatives does not matter — they commute.

Problems of Chapter 4

4.1 Thermodynamic entropy of an ideal gas

a) The integrated density of states $\Omega(E)$ of the ideal gas is given by

$$\begin{aligned}\Omega(E) &= \frac{1}{N!h^{3N}}\int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{p}_1\cdots\mathrm{d}\mathbf{p}_N \int_{V^N} \mathrm{d}\mathbf{q}_1\cdots\mathrm{d}\mathbf{q}_N\,\theta(E-H(\mathbf{q},\mathbf{p}))\\ &= \frac{1}{N!h^{3N}}V^N\int_{\mathbf{R}^{3N}} \mathrm{d}\mathbf{p}_1\cdots\mathrm{d}\mathbf{p}_N\,\theta\left(E-\frac{1}{2m}\sum_{n=1}^{N}|\mathbf{p}_n|^2\right). \qquad (12.47)\end{aligned}$$

The constant h is inserted for dimensional reasons. Now use that the volume of a hypersphere of radius $r=\sqrt{2mE}$ in dimension $3N$ equals

$$\text{volume} = \frac{\pi^{3N/2}r^{3N}}{\Gamma(z)} \qquad \text{with } z=\frac{3N}{2}+1. \qquad (12.48)$$

Hence one obtains

$$\Omega(E) = \frac{1}{N!h^{3N}}V^N\frac{1}{\Gamma(\frac{3N}{2}+1)}\left(2\pi mE\right)^{3N/2}. \qquad (12.49)$$

Taking the derivative with respect to E yields the result for $\omega(E)$.
b) By taking the logarithm one obtains

$$\mathcal{S}(E) = k_\mathrm{B}\ln\Omega(E) = k_\mathrm{B}N\left[\ln\frac{V}{h^3}+\frac{3}{2}\ln 2\pi mE-\frac{1}{N}\ln N!\Gamma(\frac{3N}{2}+1)\right]. \qquad (12.50)$$

Using Sterling's approximation there follows

$$\mathcal{S}(E) \simeq k_\mathrm{B}N\left[\ln\frac{V}{h^3}+\frac{3}{2}\ln 2\pi mE-\frac{5}{2}\ln N\right]. \qquad (12.51)$$

Finally, introduce constants ϵ and a to make the arguments of the logarithms dimensionless. The term $-\frac{5}{2}\ln N$ is split into two parts to compensate for the extensiveness of both V and E. The result can then be written as (4.40).
c) Using

$$\frac{\mathrm{d}\mathcal{S}}{\mathrm{d}E} = \frac{1}{T} \qquad (12.52)$$

as the definition of the temperature T, one finds immediately

$$\frac{1}{T} = \frac{3}{2}k_\mathrm{B}\frac{N}{E}, \qquad (12.53)$$

which is the equipartition law $E=\frac{3}{2}Nk_\mathrm{B}T$.

4.2 The quantum harmonic oscillator
One calculates

$$\begin{aligned} S(E) &= -\operatorname{Tr} \rho_E \ln \rho_E \\ &= -\sum_n e^{-|z|^2} \frac{1}{n!} |z|^{2n} \ln e^{-|z|^2} \frac{1}{n!} |z|^{2n} \\ &= |z|^2 + e^{-|z|^2} \sum_n \frac{1}{n!} |z|^{2n} \ln n! - e^{-|z|^2} \sum_n \frac{1}{n!} |z|^{2n} \ln |z|^{2n} \\ &= |z|^2 + e^{-|z|^2} \sum_n \frac{1}{n!} |z|^{2n} \ln n! - |z|^2 \ln |z|^2. \end{aligned} \tag{12.54}$$

The series appearing in this expression is convergent. Indeed, using $\ln n! \leq n(n-1)$, there follows

$$\begin{aligned} S(E) &\leq |z|^2 + e^{-|z|^2} \sum_{n=2}^{\infty} \frac{1}{n!} |z|^{2n} n(n-1) - |z|^2 \ln |z|^2 \\ &= |z|^2 + |z|^4 - |z|^2 \ln |z|^2. \end{aligned} \tag{12.55}$$

Remember that $|z|^2 = \frac{E}{\hbar\omega_0} - \frac{1}{2}$. Hence, the derivative becomes

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}E} &= \frac{1}{\hbar\omega_0} \frac{\mathrm{d}S}{\mathrm{d}|z|^2} \\ &= \frac{1}{\hbar\omega_0} \left(-\ln |z|^2 + e^{-|z|^2} \sum_{n=1}^{\infty} \frac{1}{n!} |z|^{2n} \ln(n+1) \right). \end{aligned} \tag{12.56}$$

For $|z|^2 < 1$ this expression is manifestly positive. For $|z|^2 > 1$ some extra work has to be done.

With modern technology it is easy to convince oneself that this derivative is a positive and decreasing function by making a plot of it. A more formal argument goes as follows. Introduce positive coefficients λ_n, satisfying $\sum_{n=0}^{\infty} \lambda_n = 1$, given by

$$\lambda_n = e^{-|z|^2} \frac{1}{n!} |z|^{2n}. \tag{12.57}$$

Next write

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}E} &= \frac{1}{\hbar\omega_0} \left(-\ln |z|^2 + \frac{1}{|z|^2} e^{-|z|^2} \sum_{n=1}^{\infty} \frac{1}{(n+1)!} |z|^{2(n+1)} (n+1) \ln(n+1) \right) \\ &= \frac{1}{\hbar\omega_0} \left(-\ln |z|^2 + \frac{1}{|z|^2} e^{-|z|^2} \sum_{n=2}^{\infty} \frac{1}{n!} |z|^{2n} n \ln n \right) \\ &= \frac{1}{\hbar\omega_0} \left(-\ln |z|^2 + \frac{1}{|z|^2} \sum_{n=0}^{\infty} \lambda_n n \ln n \right). \end{aligned} \tag{12.58}$$

Now use the convexity of the function $n \ln n$ to estimate

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}E} &\geq \frac{1}{\hbar\omega_0}\left(-\ln|z|^2 + \frac{1}{|z|^2}\left(\sum_{n=0}^{\infty}\lambda_n n\right)\ln\sum_{n=0}^{\infty}\lambda_n n\right)\\ &= 0.\end{aligned} \tag{12.59}$$

To obtain the latter, $\sum_{n=0}^{\infty}\lambda_n n = |z|^2$ has been used. One concludes that the first derivative of $S(E)$ cannot be negative.

The second derivative equals

$$\begin{aligned}\frac{\mathrm{d}^2S}{\mathrm{d}E^2} &= \frac{1}{(\hbar\omega_0)^2}\Bigg(-\frac{1}{|z|^2} - e^{-|z|^2}\sum_{n=1}^{\infty}\frac{1}{n!}|z|^{2n}\ln(n+1)\\ &\quad + e^{-|z|^2}\sum_{n=1}^{\infty}\frac{1}{(n-1)!}|z|^{2(n-1)}\ln(n+1)\Bigg)\\ &= \frac{1}{(\hbar\omega_0)^2}\left(-\frac{1}{|z|^2} + e^{-|z|^2}\sum_{n=0}^{\infty}\frac{1}{n!}|z|^{2n}\ln\frac{n+2}{n+1}\right)\\ &= \frac{1}{(\hbar\omega_0)^2}\frac{1}{|z|^2}\left(-1 + \sum_{n=0}^{\infty}\lambda_n n\ln\left(1+\frac{1}{n}\right) - \lambda_0\right).\end{aligned} \tag{12.60}$$

Now use that the function $n\ln(1+1/n)$ is concave to estimate

$$\begin{aligned}\frac{\mathrm{d}^2S}{\mathrm{d}E^2} &\leq \frac{1}{(\hbar\omega_0)^2}\frac{1}{|z|^2}\left[-1 + |z|^2\ln\left(1+\frac{1}{|z|^2}\right) - \lambda_0\right]\\ &\leq 0.\end{aligned} \tag{12.61}$$

The latter follows because of $x\ln(1+1/x) \leq 1 + e^{-x}$. One concludes that the second derivative is negative.

Problems of Chapter 7

7.1 The kappa-distribution

One clearly needs the identification $-(1+\kappa) = 1/(1-q)$. This can be written as $q = \dfrac{2+\kappa}{1+\kappa}$. From the identification of

$$\exp_q(-u) = \frac{1}{[1+(q-1)u]^{1/(q-1)}} = \frac{1}{[1+\frac{u}{1+\kappa}]^{1+\kappa}} \quad \text{with} \quad \frac{1}{\left[1+\frac{1}{\kappa-a}\frac{v^2}{v_0^2}\right]^{1+\kappa}} \tag{12.62}$$

then follows that $u = \frac{1+\kappa}{\kappa - a}\frac{v^2}{v_0^2}$. Hence, $\frac{m}{2\sigma^2} = \frac{1+\kappa}{\kappa - a}\frac{1}{v_0^2}$ is needed. The result is then

$$f_\kappa(v) = \frac{c(v)}{A(\kappa)v_0}\exp_q(-\frac{1}{2\sigma^2}mv^2). \tag{12.63}$$

Finally, comparison of the prefactors gives

$$A(\kappa) = d_q\left(\frac{1}{2}\frac{\kappa - a}{1+\kappa}\right)^{3/2}. \tag{12.64}$$

7.2 The Student's t-distribution
Note that

$$g_\lambda(v) = \frac{\lambda}{\sqrt{n\pi}}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}\left(1 + \frac{\lambda^2 v^2}{n}\right)^{-(n+1)/2}. \tag{12.65}$$

Identification of the exponents gives $\kappa = (1+n)/2$. From the v-dependence then follows

$$\lambda^2 = \frac{n}{\kappa - a}\frac{1}{v_0^2}. \tag{12.66}$$

Make the *ansatz* that $a = 1/2$. Then one obtains $\lambda^2 v_0^2 = 2$. Finally one can identify the prefactors. This yields

$$\frac{c(v)}{A(\kappa)} = \sqrt{\frac{2}{n\pi}}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}. \tag{12.67}$$

7.3 Order statistics
Solution of the $q < 1$-case
The probability that $u < u_0$ is

$$\begin{aligned} G(u_0) &= \text{Prob}(u_1 < u_0) + \text{Prob}(u_1 > u_0)\,\text{Prob}(u_2 < u_0) + \cdots \\ &= \sum_{k=1}^{n}\left(1 - \frac{u_0}{(n-1)T}\right)^{k-1}\frac{u_0}{(n-1)T} \\ &= 1 - \left(1 - \frac{u_0}{(n-1)T}\right)^{n}. \end{aligned} \tag{12.68}$$

By taking the derivative one finds

$$p_T(u) = \frac{\mathrm{d}}{\mathrm{d}u}G(u) = \frac{n}{n-1}T\left(1 - \frac{u_0}{(n-1)T}\right)^{n-1}. \tag{12.69}$$

This can be written into the form

$$p_T(u) = \left(1 - (1-q)\Phi(\theta) - (1-q)\theta x\right)^{1/(1-q)}, \tag{12.70}$$

with $q = (n-2)/(n-1)$ and

$$\theta = (2-q)^{1-q} T^{-q}. \tag{12.71}$$

Solution of the $q > 1$-case
First calculate

$$F(u) = \int_0^u \mathrm{d}v\, f(v) = \frac{cu}{1+cu}. \tag{12.72}$$

The probability that $u < u_0$ is

$$\begin{aligned} G(u_0) &= \sum_{k=1}^{n} \left(1 - F(u)\right)^{k-1} F(u) \\ &= 1 - \left(1 - F(u)\right)^n. \end{aligned} \tag{12.73}$$

By taking the derivative one finds

$$p_T(u) = n(1-F(u))^{n-1} f(u) = \frac{nc}{(1+cu)^{n+1}}. \tag{12.74}$$

This can be written into the form

$$p_T(u) = \left(\frac{1}{1 + (q-1)\Phi(\theta) + (q-1)\theta x}\right)^{1/(q-1)}, \tag{12.75}$$

with $q = (n+2)/(n+1)$ and

$$\theta = \frac{1}{(q-1)n^{q-1}} c^{2-q} = \left(\frac{n+1}{n}\right)^{1/(n+1)} T^{-n/(n+1)}. \tag{12.76}$$

7.4 Stationary solutions
Let us calculate

$$\begin{aligned} & -\frac{\partial V}{\partial x} p + \frac{\partial}{\partial x}(\lambda + \mu V) p \\ &= -V'(x)p + \mu V'(x) p + (\lambda + \mu V)\frac{\partial p}{\partial x} \\ &= (\mu - 1)V'(x)p(x) + (\lambda + \mu V)p^q(x)\left[-\theta V'(x)\right] \\ &= \left[\mu - 1 - \theta(\lambda + \mu V)p^{q-1}(x)\right] V'(x)p(x) \\ &= \left[\mu - 1 - \theta \frac{\lambda + \mu V}{1 + (q-1)\alpha(\theta) + (q-1)\theta V(x)}\right] V'(x)p(x). \end{aligned} \tag{12.77}$$

This vanishes when

$$\lambda = -\frac{1}{(2-q)\theta}(1+(q-1)\alpha(\theta))$$
$$\mu = -\frac{q-1}{2-q}. \tag{12.78}$$

7.5 One has

$$|\psi_t(x)|^2 = A^2\left(1+\frac{\omega^2 x^2}{c^2}\right)^{-\frac{1}{2}-\frac{mc^2}{\hbar\omega}}, \tag{12.79}$$

Compare this with

$$f(x) = \frac{1}{c_q\sigma}\exp_q\left(-\frac{x^2}{\sigma^2}\right) = \frac{1}{c_q\sigma}\left[1+(q-1)\frac{x^2}{\sigma^2}\right]^{-1/(q-1)}. \tag{12.80}$$

One concludes that

$$\frac{1}{q-1} = \frac{1}{2}+\frac{mc^2}{\hbar\omega} \tag{12.81}$$

and

$$(q-1)\frac{x^2}{\sigma^2} = \frac{\omega^2 x^2}{c^2}. \tag{12.82}$$

The latter implies

$$\frac{1}{\sigma^2} = \left(\frac{1}{2}+\frac{mc^2}{\hbar\omega}\right)\frac{\omega^2}{c^2}. \tag{12.83}$$

7.6 Second moment of the q-Gaussian
a) From (7.42) and (7.26) follows

$$-\langle\langle x\rangle\rangle_\theta = \frac{\mathrm{d}\alpha}{\mathrm{d}\theta} = \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\sigma^{q-1}-1}{q-1}$$
$$= \frac{1}{q-1}\frac{\mathrm{d}}{\mathrm{d}\theta}\theta^{\frac{q-1}{q-3}}$$
$$= -\frac{1}{3-q}\theta^{-\frac{2}{3-q}}. \tag{12.84}$$

In the limit $q=1$ this becomes $\langle x^2\rangle = 1/2\theta = \frac{1}{2}\sigma^2$, as it should be.
b) From (7.42) and (7.36) follows

$$S(U) = \frac{1}{2-q}\left[\alpha(\theta)+\theta U\right] = \frac{1}{2-q}\left[\frac{\theta^{\frac{q-1}{q-3}}-1}{q-1}+\theta U\right]. \tag{12.85}$$

Hence one has

$$\Phi(\theta) = S(U) - \theta U = \frac{\theta^{\frac{q-1}{q-3}} - 1}{(q-1)(2-q)} - \frac{1-q}{2-q}\theta U. \tag{12.86}$$

Taking the derivative w.r.t. θ gives

$$-U = \frac{\mathrm{d}\Phi}{\mathrm{d}\theta} = -\frac{1}{(3-q)(2-q)}\theta^{-2/(3-q)} - \frac{1-q}{2-q}U - \frac{1-q}{2-q}\theta\frac{\mathrm{d}U}{\mathrm{d}\theta}. \tag{12.87}$$

This can be written as (7.88). Note that in the limit $q = 1$ one obtains again $U = 1/2\theta$.

Problems of Chapter 8

8.1 Conventional energy constraints

Let us first verify that the pdf $p_\theta(j)$ is an extremum of the function of Lagrange

$$\mathcal{L} = S_q^{\mathrm{Tsallis}}(p) - \tilde{\alpha}\sum_j p(j) - \tilde{\theta}^k \sum_j p(j)H_k(j). \tag{12.88}$$

Taking the derivative w.r.t. $p(j)$ gives the condition

$$0 = \frac{q}{1-q}p(j)^{q-1} - \tilde{\alpha} - \tilde{\theta}^k H_k(j). \tag{12.89}$$

There follows

$$p(j)^{q-1} = \frac{1-q}{q}\left(\tilde{\alpha} + \tilde{\theta}^k H_k(j)\right). \tag{12.90}$$

This can be written as (8.68), with $\alpha(\theta) = \frac{1}{q}\tilde{\alpha}$ and $\theta^k = \frac{1-q}{q}\tilde{\theta}^k$.

Next, let us show that this extremum corresponds with an absolute maximum. Because of $\exp_q^*(u) = exp_{2-q}(u)$ the pdf $p_\theta(j)$ belongs to the $(2-q)$-exponential family. It therefore maximizes the entropy function $I(p)$, given by (7.34), with q replaced by $2-q$. But the latter equals the Tsallis entropy $S_q^{\mathrm{Tsallis}}(p)$. Therefore, the extremum realised by $p_\theta(j)$ corresponds with an absolute maximum within the set of all pdfs giving the same expectations for the energies $\langle H_k \rangle_p$.

8.2 Identity

From the definitions of the q-exponential family and of Tsallis' entropy one obtains

$$\exp_q\left[S_q^{\mathrm{Tsallis}}(p)\right] = \exp_q\left[\frac{1}{1-q}\left(\sum_j p(j)^q - 1\right)\right]$$

$$= \left[1 + \left(\sum_j p(j)^q - 1\right)\right]^{1/(1-q)}$$
$$= \left[\sum_j p(j)^q\right]^{1/(1-q)}$$
$$= z^{1/(1-q)} \quad \text{with} \quad z = \sum_j p(j)^q. \tag{12.91}$$

Substituting p by P and q by $1/q$ this yields

$$\exp_{1/q}\left[S_{1/q}^{\text{Tsallis}}(P)\right] = \left[\sum_j P(j)^{1/q}\right]^{1/(1-1/q)}. \tag{12.92}$$

Using $P(j) = p(j)/z$ and $\sum_j p(j) = 1$ the latter becomes

$$\exp_{1/q}\left[S_{1/q}^{\text{Tsallis}}(P)\right] = \left[\sum_j z^{-1/q} p(j)\right]^{1/(1-1/q)} = z^{1/(1-q)}. \tag{12.93}$$

This shows that (12.91) and (12.93) are equal.

Problems of Chapter 9

9.1 Equivalent entropies

Use that $\sum_j p(j)^q = 1 + (1-q) S_q^{\text{Tsallis}}(p)$ to write

$$S_{qr}^{\text{SM}}(p) = \frac{1}{1-r}\left(\left[1 + (1-q) S_q^{\text{Tsallis}}(p)\right]^{(r-1)/(q-1)} - 1\right). \tag{12.94}$$

Hence, it is obvious to define the function $\xi(u)$ by

$$\xi(u) = \frac{1}{1-r}\left([1 + (1-q)u]^{(r-1)/(q-1)} - 1\right). \tag{12.95}$$

Its first derivative equals

$$\xi'(u) = [1 + (1-q)x]^{\frac{r-1}{q-1} - 1} \tag{12.96}$$

and is always positive. Hence, $S_{qr}^{\text{SM}}(p)$ is a monotonically increasing function of $S_q^{\text{Tsallis}}(p)$.

Problems of Chapter 10

10.1 Heine's distribution

This is a simple Problem once one knows the answer. Let us assume that we do not know it. Taking the derivative w.r.t. λ gives

$$\frac{\partial}{\partial\lambda}p_\lambda(n) = p_\lambda(n)\left[\frac{n}{\lambda} - (1-q)\sum_{j=1}^{\infty}\frac{q^{j-1}}{1+\lambda(1-q)q^{j-1}}\right]. \tag{12.97}$$

Hence one has, after multiplication with $-\lambda$,

$$-\lambda\frac{\partial}{\partial\lambda}p_\lambda(n) = p_\lambda(n)\left[-n + (1-q)\lambda\sum_{j=1}^{\infty}\frac{q^{j-1}}{1+\lambda(1-q)q^{j-1}}\right]. \tag{12.98}$$

Note that within the square brackets a separation has been obtained between the variable n and the parameter dependence. This shows that the escort probability of $p_\lambda(n)$ is $p_\lambda(n)$ itself, that the appropriate parameter is $\theta = -\ln\lambda$, and that the corresponding Hamiltonian is $H(n) = n$. The probability distribution therefore belongs to the exponential family. However, after integration of (12.98) the integration constant depends on the variable n. It is therefore necessary to introduce a second parameter θ^2, which in the case of the Heine distribution equals 1.

Now let $\theta^1 \equiv \theta = -\ln\lambda$ and $H_1(n) \equiv H(n) = n$. Taking the logarithm of (12.98) gives

$$\ln p_\lambda(n) = = -n\theta^1 + \ln e_q(-\lambda) + \frac{1}{2}n(n-1)\ln q - ln[n]_q!. \tag{12.99}$$

Introduce therefore

$$H_2(n) = -\frac{1}{2}n(n-1)\ln q + ln[n]_q!. \tag{12.100}$$

Then one can has $p_\lambda(n) = p_\theta(n)$

$$p_\theta(n) = \exp(-\Phi(\theta) - \theta^k H_k(n), \tag{12.101}$$

with $\theta^1 = \lambda$, $\theta^2 = 1$, and

$$\Phi(\theta) = \ln\sum_{n=0}^{\infty} e^{-\theta^k H_k(n)}. \tag{12.102}$$

10.2 Lambert's W function

From $W(0) = 0$ follows that $\ln_W(1) = 0$. From $W(-1/e) = -1$ follows $\ln_W(0) = -1$. The derivative is

$$\frac{\mathrm{d}}{\mathrm{d}u}\ln_W(u) = \frac{\mathrm{d}}{\mathrm{d}u}W\left(\frac{u-1}{e}\right) = \frac{1}{u-1}\frac{\ln_W(u)}{1+\ln_W(u)} > 0. \quad (12.103)$$

This means that $\ln_W(u)$ is a strictly increasing function on the positive real axis. From the series expansion

$$W(u) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} u^n \quad (12.104)$$

follows

$$\ln_W(u) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} \frac{(u-1)^n}{e^n}. \quad (12.105)$$

In particular, the singularity at $u = 1$ in (12.103) is only apparent. One has

$$\frac{\mathrm{d}}{\mathrm{d}u}\ln_W(u)\bigg|_{u=1} = \frac{1}{e}. \quad (12.106)$$

The second derivative is

$$\frac{\mathrm{d}^2}{\mathrm{d}u^2}\ln_W(u) = \frac{-1}{(u-1)^2}\,\frac{\ln_W(u)^2(2+\ln_W(u))}{(1+\ln_W(u))^3} < 0. \quad (12.107)$$

Hence the function $\ln_W(u)$ is concave.

The range of $\ln_W(u)$ is the interval $[-1,+\infty)$, with $\ln_W(0) = -1$. Hence, the inverse function $\exp_W(v)$ is put equal to 0 on the interval $(-\infty,-1]$. On the interval $[-1,+\infty)$ the inverse function $\exp_W(v)$ is obtained by solving $\ln_W(u) = v \Leftrightarrow u = \exp_W(v)$. But from

$$\begin{aligned} v = \ln_W(u) &= W(\frac{1}{e}(u-1)) = \frac{1}{e}(u-1)\exp(-W(\frac{1}{e}(u-1))) \\ &= \frac{1}{e}(u-1)e^{-v} \end{aligned} \quad (12.108)$$

follows

$$\exp_W(v) = 1 + ve^{v+1} \qquad \text{if } v \geq -1. \quad (12.109)$$

Problems of Chapter 11

11.1 q-deformed relative entropy

The deformed logarithm equals

$$\ln_\phi(u) = \int_1^u \mathrm{d}v\,\frac{1}{\phi(v)} = \int_1^u \mathrm{d}v\,v^{-q} = \frac{1}{1-q}\left(u^{1-q}-1\right). \quad (12.110)$$

Hence, (11.9) becomes

$$\begin{aligned} D_\phi(p||r) &= s\sum_j \int_{r(j)}^{p(j)} \mathrm{d}v\ [\ln_\phi(v) - \ln_\phi(r(j))] \\ &= \frac{2-q}{1-q}\sum_j \int_{r(j)}^{p(j)} \mathrm{d}v\ \left[v^{1-q} - r((j)^{1-q}\right] \\ &= \frac{1}{1-q}\sum_j \left[p(j)^{2-q} - r(j)^{2-q}\right] - \frac{2-q}{1-q}\sum_j [p(j)-r(j)]\,r(j)^{1-q} \\ &= -S_{2-q}^{\text{Tsallis}}(p) + S_{2-q}^{\text{Tsallis}}(r) - \frac{2-q}{1-q}\sum_j [p(j)-r(j)]\,r(j)^{1-q}. \end{aligned} \tag{12.111}$$

The latter is (8.49).

11.2 A two-parameter family
Comparison of (11.54) with (11.1) yields $s = 1$ and

$$\omega_\phi(u) = \frac{1}{q-r}\left(u^{1-r} - u^{1-q}\right). \tag{12.112}$$

Hence one has,

$$u\omega_\phi(\frac{1}{u}) = \frac{1}{q-r}\left(u^r - u^q\right). \tag{12.113}$$

Taking the derivative, using (10.25), one obtains

$$\begin{aligned} \ln_\phi(u) + \text{ constant } &= -\frac{\mathrm{d}}{\mathrm{d}u}u\omega_\phi\left(\frac{1}{u}\right) \\ &= \frac{q}{q-r}u^{q-1} - \frac{r}{q-r}u^{r-1}. \end{aligned} \tag{12.114}$$

Taking a further derivative gives

$$\frac{1}{\phi(u)} = \frac{q(q-1)}{q-r}u^{q-2} - \frac{r(r-1)}{q-r}u^{r-2}. \tag{12.115}$$

This can be written as

$$\phi(u) = \frac{(q-r)u^2}{q(q-1)u^q - r(r-1)u^r}. \tag{12.116}$$

The corresponding deformed logarithm is

$$\ln_\phi(u) = \int_1^u \mathrm{d}v\,\frac{1}{\phi(v)} = \frac{1}{q-r}\int_1^u \mathrm{d}v\,(q(q-1)u^{q-2} - r(r-1)u^{r-2})$$

$$= \frac{1}{q-r}(qu^{q-1} - ru^{r-1}) - 1. \tag{12.117}$$

11.3 Kaniadakis' entropy functional

From the previous Problem, using $q = 1 + \kappa$ and $r = 1 - \kappa$, there follows

$$\phi(u) = \frac{2u}{(1+\kappa)u^{\kappa} + (1-\kappa)u^{-\kappa}}. \tag{12.118}$$

It is straightforward to verify that $u\phi'(u) = (1-\kappa)\phi(u)$ holds.

Index

additivity, 134
alpha entropies, 132

BGS-entropy functional, 39
bias, 23
binomial distribution, 35
Bloch representation, 24, 50
Boltzmann entropy, 56
Boltzmann's constant, 3, 18, 31
Boltzmann-Gibbs distribution, 5, 71
Boltzmann-Gibbs-Shannon entropy functional, 70
Boltzmann-Gibbs-von Neumann density operator, 14
Borges-Roditi entropy functional, 128
Bose-Einstein distribution, 18
box dimension, 137
Bregman type divergence, 124, 166

canonical ensemble, 5, 53
Cantor set, 139
Cauchy distribution, 27, 156
central limit theorem, vi
chemical potential, 10
Clausius' notion of entropy, 38
coarse graining, 137
coding function, 170
coherent states, 91
completely monotonic, 72
complexity, 169
concave function, 39, 42, 48, 95
configurational partition sum, 7
configurational pdf, 106
configurational temperature, 136
conserved quantities, 54
contact transform, 42, 77
control parameter, 10
convex function, 40, 95
corrector, 170
correlation function, 34
Cramér and Rao inequality, 172
curved exponential family, 28
curved quantum exponential family, 34

deduced logarithm, 151, 165, 177
deformation index, 98
deformed exponential function, v, 95, 149
deformed logarithmic function, v, 95, 149
density matrix, 12
density of states, 31, 55, 56, 65, 106
density operator, 12, 23
divergence, 45, 51, 123, 124, 166
dual identities, 43, 168
dynamical entropy, 89

effective temperature, 129
Einstein convention, 26
entanglement, 62
entropy, 38, 42
entropy encoding, 39
entropy functional, 38, 39
equilibrium distribution, 42
equilibrium state, 37, 38
equipartition, 4, 7, 12, 56
equivalence of ensembles, vi, 67
ergodic hypothesis, 54
ergodic theorem, 54, 66
escort density operators, 109, 160
escort probability distribution, v, 100, 102, 138, 153, 172
exact differential, 46
exponential distribution, 35, 48

J. Naudts, *Generalised Thermostatistics*, DOI 10.1007/978-0-85729-355-8,

exponential family, 21, 26, 140
extensive variables, 10

f-divergence, 123
Fermi-Dirac distribution, 18
Fibonacci numbers, 139
Fisher information, 172
fractal, 137
free energy, 43, 138

game theory, 169
Gamma distribution, 34
Gauss distribution, 28
generalised dimension, 137
generalised Lorentzian, 110
Gibbs distribution, 5, 71
Gibbs' principle of minimal free energy, 45
grand-canonical ensemble, 10
grand-canonical partition sum, 11

Hamilton's equations of motion, 6
harmonic oscillator, 16, 32, 55, 73
Hausdorff dimension, 137
heat bath, 10
heat capacity, 15, 46
heat capacity, negative, 57
Heine's distribution, 161
hyperdistribution, 69, 74
hyperensemble, 74, 77

ideal Bose gas, 17, 33
ideal Fermi gas, 17, 33
ideal gas, 7, 12, 18, 65
ideal gas law, 12
iid, 79, 157
information matrix, 172
intensive variable, 10
internal energy, 15, 46
inverse temperature, 44, 77
Ising model, 21, 29, 43, 58, 81
isolated, 53

Jackson's q-deformed exponential function, 161

Kaniadakis' deformed logarithm, 152
kappa-distribution, 110
Kolmogorov-Nagumo average, 131
Kullback-Leibler distance, 45, 123

Lagrange optimisation, 41
Lagrange parameter, 70
large deviation theory, vi
lattice animals, 158
law of Dulong and Petit, 64
Legendre transform, 42, 77
linear entropy, 116
Lorentzian, 27

Markov chain, 86
Massieu function, 23, 42, 75, 138, 168
maximum entropy principle, 38, 48, 70, 74
maximum likelihood, 27
Maxwell distribution, 3, 105
Maxwell relation, 50
mean field equation, 82
method of Lagrange, 41
metric tensor, 171
microcanonical ensemble, 42, 53, 56, 105
mono-atomic gas, 106
multifractal analysis, 137

negative heat capacity, 57
noise, 55
non-additive entropy, 117, 134
non-equilibrium Massieu function, 45
non-equilibrium statistical physics, vi
non-extensive thermostatistics, 117
normal distribution, 28
number of degrees of freedom, 111

open boundary conditions, 29

partition sum, 5, 26
Pauli matrices, 19, 24, 63
Pauli spin, 50
pendulum, 59
percolation, 158
periodic boundary conditions, 30
phase separation, 58
phase space, 4, 32, 54
phase transition, 58, 82
phi-deformed entropy functional, 165
phi-exponential family, 153
Poisson distribution, 26
positive operator-valued measure, 77
potential, 23
POVM, 77
pressure, 12
prior probability, 26
prior weight, 26
probability distribution in phase space, 4
product measure, 79

q-exponential family, 98

q-Gaussian distribution, 103
q-Maxwellian, 105
quantum exponential family, 32
quantum microcanonical ensemble, 62
quantum phi-exponential family, 160
quantum q-exponential family, 109
quantum variational principle, 49
quasi-stationary process, 46
quasifree states, 90

Rényi dimension, 137
Rényi's entropy functional, 132
Raleigh distribution, 35
real gas, 106
relative complexity, 169
relative entropy, 45, 48, 123
robustness, 170

Sackur-Tetrode equation, 66
scaling, 137
scaling function, 131
self-dual, 151, 152
Sharma-Mittal entropy functional, 143
source coding theorem, 39
stability, 37, 46, 58, 80, 82, 123, 126, 170
stationary state, 88
statistical manifold, 177
Stirling's approximation, 142
stochastic variable, 23
Student's t-distribution, 111
superstatistical distribution, 73
superstatistics, 77, 99
susceptibility, 31

temperature, 44, 77, 135
theorem of Liouville, 6
thermodynamic entropy, 38, 42, 76, 138
thermodynamic equilibrium, 37
thermodynamic forces, 126
thermodynamic limit, v, vi, 30, 67, 137
thermodynamic relation, 43, 76
thermodynamic stability, 37, 46
thermodynamic variables, 37
time average, 54
trace class operator, 47
transfer matrix method, 30, 83
Tsallis distribution, 118, 119
Tsallis' entropy functional, 50, 115, 133
Tsallis' thermostatistics, v
two-level atom, 63

unbiased estimator, 23

variational principle, 45, 140, 167, 174
von Neumann density operator, 14
von Neumann entropy, 47, 50, 63

wave function, 13
Weibull distribution, 35
weight, 26, 39