Publishers' page

Publishers' page

Publishers' page

Publishers' page

Generalized Statistics and its Applications

Petr Jizba

Habilitation thesis

Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering
Department of Physics

Prague, November 2021

To My Family

Hana, Anežka, Klára and Jan

Preface

During the past two decades, the Boltzmann–Gibbs statistical mechanics (BGS) undergone an important conceptual shift. An important catalyst has been the massive infusion of new ideas from the theory of critical phenomena (scaling laws, renormalization group, etc.), (multi-) fractals and trees, random matrix theory, network theory, and non-Shannonian information theory. While successful in describing stationary systems characterized by ergodicity or metric transitivity, BGS fails to reproduce statistical behaviour of many real-world systems encountered, e.g., in biology, astrophysics, geology, climatology, economics or social sciences. Typical signatures of the BGS are Exponential and Gaussian distributions. These distributions are those that maximise the Boltzmann–Gibbs or Shannon entropy and ensure the equilibrium state. The Maxwell distribution is an instance of the ensuing equilibrium distribution for the velocities of molecules in an ideal gas. The underlying mathematical reason for this can be expressed in terms of the standard Central Limit Theorem (CLT) and its various information-theoretic reincarnations. On the other hand, ergodicity breakdown, strong correlations, long-range interactions or non-equilibrium conditions, such as energy flows through dissipative systems, are all mechanisms that can push a system away from the typical equilibrium states otherwise predicted by BGS.

Recently the use of a new paradigm known as *generalized statistics* has become very popular especially in physics and biology. In this connection, the field of complex dynamical systems has provided a fertile ground for a systematic study of various generalized-statistics frameworks, which are pertinent to large classes of systems that do not conform to BGS. The notion of generalized statistics refers to statistical systems that are characterized by fat- (or heavy-) tailed rather than Exponential-type distribu-

tions. Fat-tailed distributions concentrate much statistical weight in the tails of the distribution and consequently rare events occur more frequently than a naive use of the BGS would indicate. Fat-tailed distribution functions are ubiquitous in complex systems and processes. Typical arenas are in nonlinear dynamics, network theory, chemistry, climatology, social sciences and finance. Examples of relevant fat-tailed distributions include Log-normal distribution, Weibull distributions, Fréchet distribution, Stretched exponential distributions, and various power-law tail distributions (such as Zipf–Pareto, Lévy, or q-exponential distributions). Especially, the statistics associated with power-law tails accounts for a rich class of phenomena that have been observed in numerous experimental, observational and model systems. Impressive experimental examples include; velocity differences measured in a Couette–Taylor experiment for a fully developed turbulence regime, transverse momentum spectra of hadrons at the Large Hadron Collider experiments in CERN, transport properties of cold atoms in dissipative optical lattices, and confined granular matter. As for observational systems, I can mention, e.g., rotation periods and diameters of asteroids and distribution of meteor showers, rotation curves of galaxies, citation of scientific papers, magnitude of earthquakes and population of cities. Among model systems, one of the paradigmatic dissipative low dimensional models, the logistic map, exhibits power-law tail behavior in the vicinity of the chaos threshold.

The purpose of this thesis is to offer my personal viewpoint on the subject of generalized statistics. In particular, the route to the generalized statistics that I wish to emphasize here hinges on such concepts as information theory of Rényi, q-deformed thermostatistics of Tsallis, Superstatistics of Beck *et al.*, and various generalized non-extensive entropies. Respective topics are introduced and logically knitted in three chapters. While Chapter 1 serves as an introductory chapter where some essentials about the generalized statistics are introduced and discussed, Chapter 2 and Chapter 3 deal with more specific issues that represent recurring themes in my research. Notably, in Chapter 2 I outline and discuss generalized statistics based on the concept of Rényi entropy, whereas in Chapter 3 selected topics in generalized statistics are elucidated from the point of view of superstatistics. Further, each chapter is supplied with four reprinted articles of mine that are logically connected with the topic of each chapter. The thesis is also accompanied with "Glossary of relevant financial terms", "Some important biographies" and a number of technical appendices where more specialized issues related to the bulk of the thesis are presented and proved.

I would like to thank a number of my colleagues and students for discussions, and/or their comments on parts of this habilitation thesis. I would also like to thank my close collaborators for shearing with me their wisdom, expertise and stimulating friendly atmosphere during writing our mutual papers — some of which are reprinted with their consent in this work. My thanks go especially to Toshihico Arimitsu, Dorje Brody, Hideki Matsumoto, Hagen Kleinert, Sumioshi Abe, Peter Harremös, Jan Korbel, Václav Zatloukal, Ray Rivers, Flemming Topsøe, Tamás Biro, Peter Van, Stefan Thurner, Christian Beck, Jacob Dunningham, Zbigniew Haba, Fabio Scardigli and Peter Haener. I am also grateful to Massimo Blasone for his assistance with the World Scientific LaTeX macro packages.

Finally I am greatly indebted to my wife Hana and my children Anežka, Klára and Jan for their moral support and patience.

Prague,
November 2021

Petr Jizba

Contents

Chapter 1

Going Beyond Boltzmann–Gibbs Statistics

The Boltzmann–Gibbs entropy and its associated statistical mechanics are producing countless contributions to our knowledge about systems whose collective dynamics satisfies simplifying hypothesis such as ergodicity, metric transitivity, and independence or quasi-independence of relevant random variables. Nevertheless, a fascinating world exists outside these assumptions. In this chapter I discuss the conceptual rationale that justify existence of this world. The corresponding mathematical underpinning will be located in the Generalized Central Limit Theorem. The passage to physics will be established via the concept of information-theoretic entropies. Various concepts of entropy are presented. Among them, I pay a particular attentions to the non-extensive entropy of Tsallis–Havrda–Charvát and extensive entropy of Rényi.

1.1 Breakdown of Boltzmann–Gibbs statistics

The standpoint adopted throughout this thesis is that statistical physics is mathematical statistics applied to physics. Maxwell, Boltzmann and (foremost) Gibbs should not be blamed for not having adopted this point of view. They established field of statistical physics more than 30 years before Kolmogorov formulated the fundamentals of mathematical statistics, more than 40 years before Cramér laid out the mathematical theory of large deviations and more then 70 years before first versions of the Lévy–Gnedenko's Generalized Central Limit Theorem appeared. The essence of statistical physics is clearly a model building, and as such it only approximates the physical reality. It consists of a probability distribution, or its quantum-mechanical analogue — density matrix, which is used to calculate statistical averages. It depends on a small number of the so-called

state parameters that can be estimated from experimental data. The roots of statistical physics lie in thermodynamics, a nineteenth century science predating statistical physics by almost 50 years. Nowadays it is a common practice to explain thermodynamical concepts in terms of statistical physics, rather than the other way round. There is yet another conceptual route to statistical physics, namely the route based on information theory. This route has revolved around the concept of information entropy found by Shannon at the end of 1940's and in statistical physics popularized by Jaynes in 1970's and 1980's. Here I will follow both routes with some bias towards information-theoretic approaches.

In recent years, a considerable effort has been made to extend the statistical-physics paradigm beyond the limits set out by Gibbs in his 1902 book "Elementary Principles of Statistical Mechanics". Traditional statistical physics focuses on systems with many degrees of freedoms and accommodates such simplifying assumptions as ergodicity, or (quasi-) independence. The formalism becomes exact in the thermodynamic limit, i.e. in the limit of infinitely many degrees of freedom. In such a case the so-called extensive thermodynamic variables acquire their true extensive status (e.g., surface effects are irrelevant). One natural motivation to go beyond the standard formalism is therefore the current interest in relatively small systems, such as Bose–Einstein condensates, single macromolecules, molecular clusters or molecular engines. Further motivation is to consider strongly correlated degrees of freedom, which can be found, for instance, in strongly correlated electron systems, in systems with long-time memory, in systems with long-range interactions or in multipartite entangled systems. Many new insights have originated from the paradigm known as *generalized statistics* which tries to justify a plethoric appearance of fat-tailed distributions in natural and social sciences. These include, Tsallis' thermostatistics [311, 312], Kaniadakis' κ entropy [162, 163], Beck's *et al.* Superstatistics [23, 24, 333] or Rényi's extensive entropy [144, 263, 264]. Despite the fact that these domains of research are not older than 25 years, they have already established an impressive publication record. For instance, Tsallis' thermostatistics paper alone has yielded to date more than 5000 publications.

1.1.1 *Central limit theorem*

A lot of important phenomena addressed by statistical physics can be conveniently described as the compound effect of many small random influences. The observable effects can be then attributed to the sum of a very large

number of such events. Paradigmatic example of this is the pressure that an ideal gas exerts on the walls of the containment vessel. Pressure as such is a state variable resulting from a large number of collisions of gas molecules with the walls. Already in 1730's D. Bernoulli realized that in order to evaluate the pressure one needs an average effect of the sum of impulses that the gas molecules impart on the walls when colliding with and recoiling from them. Simple kinematic reasonings then led him to the formula (see, e.g., Ref. [187]):

$$p \;=\; \frac{1}{Vm}\sum_{i=1}^{N}\frac{p_{ix}^2+p_{iy}^2+p_{iz}^2}{3} \;=\; \frac{2}{3V}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m} \;=\; \frac{2}{3}\,n\langle\varepsilon\rangle\,. \qquad (1.1)$$

The latter is presently known as Bernoulli's formula for pressure. Here V is the vessel volume, N is the total number of molecules, n is the number of molecules per unit volume and m is the corresponding molecule mass. In his own words, by "$\langle\varepsilon\rangle$" Bernoulli meant [33] "...the average value of kinetic energy of a single molecule", but what he actually mathematically done was a simple arithmetical average of the kinetic energy of all participating molecules. One may thus ask, in what sense the arithmetical average of kinetic energies could be equated with the average kinetic energy of a *single* particle, particularly when the underlying single-particle momentum distribution could not be known — certainly not in Bernoulli's time. In fact it took another 140 years before Maxwell–Boltzmann distribution was discovered and both predictions could be compared.

The answer to this question provides *central limit theorem* (CLT) which was mathematically properly formulated in 1930's but it had had its inceptive already in works of A. de Moivre, P.S. Laplace or A. Lyapunov. Roughly speaking, the CLT states that the (arithmetic) mean value of a sufficiently large number (say N) of independent, identically distributed random variables, each with *finite mean* μ and *variance* σ^2, will be approximately normally (or Gaussian) distributed with parameters μ and σ^2/N.

With the development of probability in 1930's many variants of the CLT were formulated. For our purpose the so-called Lindeberg version of the CLT will be sufficient (cf. Ref. [85]):

Theorem 1.1. *Let $\{X_k\}$ be a sequence of mutually independent random variables with a common distribution. Suppose that X_k (for all k) have the finite expected value μ and finite variance σ^2 and let $S_N = X_1 + X_2 + \cdots + X_N$. Define the random variable $\hat{S}_N = (S_N - N\mu)/\sqrt{N}$, then $\hat{S}_N$ has a*

density function $f_N(x)$ which satisfies the relation

$$\lim_{N\to\infty} f_N(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right] \quad \textit{for all real numbers } x\,,$$

and the convergence is uniform in the parameter x. If a sequence of distribution functions $P_N(x),\ N = 1, 2, \ldots$ have density functions $f_N(x)$, and they satisfy previous formula, then

$$\lim_{N\to\infty} P_N(\hat{S}_N < x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left[-\frac{u^2}{2\sigma^2}\right] du\,.$$

The convergence is uniform with respect to the variable x. So in particular, S_N is in the large-N limit distributed according to normal distribution where the mean and variance are 0 and σ^2, respectively.

If, in addition, X_k (for all k) have also finite 3-rd moment then

$$f_N(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right] \;+\; \mathcal{O}\left(\frac{1}{\sqrt{N}}\right).$$

□

In Appendix B I present a simple proof of this theorem and also list some other frequently used versions of the CLT. Note that by defining the arithmetic (or sample) average $\bar{X}_N = S_N/N$, then the CLT implies

$$\lim_{N\to\infty} P_N\left(-x < \frac{\bar{X}_N - \mu}{\sigma/\sqrt{N}} < x\right) \;=\; \mathrm{erf}(x)\,, \tag{1.2}$$

where $\mathrm{erf}(x)$ is the *Gauss error function*. This, in turn, gives that for $N \gg 1$

$$P_N(-x < \bar{X}_N - \mu < x) \;\simeq\; \mathrm{erf}\left(\frac{x\sqrt{N}}{\sigma}\right) \;=\; 1 \;-\; \mathrm{erfc}\left(\frac{x\sqrt{N}}{\sigma}\right)$$

$$=\; 1 \;-\; \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{1.3}$$

Here $\mathrm{erf}(x)$ is the *complementary error function*, i.e. $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$. So, for arbitrarily small x and $N \gg 1$, the arithmetic mean *converges in probability* to μ, i.e.

$$\bar{X}_N \;\xrightarrow{P}\; \mu\,. \tag{1.4}$$

Thus, the CLT automatically subsumes a weak law of large numbers.

By going back to Bernoulli's pressure formula, I can now identify X_k with p_{kx}^2 and assume that the variance of p_{kx}^2 is finite then in the large-N limit

$$\sum_{i=1}^{N} \frac{p_{ix}^2}{N} \;\xrightarrow{P}\; \langle p_{kx}^2 \rangle\,, \tag{1.5}$$

where $\langle \cdots \rangle$ represents the average value with respect to the single-particle distribution (whatever this might be). So, in particular, for a large number of particles I can formally write

$$\frac{1}{2m} \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{N} \;=\; \left\langle \frac{\mathbf{p}_k^2}{2m} \right\rangle \;=\; \langle \varepsilon \rangle \,. \tag{1.6}$$

This last result, in turn, justifies a posteriori the correctness of Bernoulli's pressure formula (1.1). Let me emphasize, in passing, that historically the Bernoulli formula represented a major achievement because it allowed microscopically explain Boyle's empirical law for ideal gases, i.e. that the pressure is inverse proportional to the volume. Incidentally, because the Gaussian distribution in the CLT is a *stable distribution* (see Section 1.1.3 and Appendix B.6) the natural candidate for a single-particle momentum distribution is also Gaussian distribution, namely Maxwell–Boltzmann distribution.

At this stage I should particularly stress that CLT is a very fundamental concept in statistical physics. Apart from the well known fact that the measurement error (and in particular, instrumental error) in an experiment is usually well fitted with a Gaussian distribution — evidently because an observed error emerges from a sum of many random (typically) independent error influences, the CLT allows rigorously justify an equivalence between *microcanonical* and *canonical ensembles*.

In statistical physics the problem of the equivalence of ensembles goes back to Boltzmann and Gibbs. In addition, the ensemble equivalence is not confined to statistical physics only; it can be found in other areas of applied probability theory, for instance, in information theory. To my knowledge the first who use the CLT to prove the ensemble equivalence was A.Y. Khintchine [168, 170] who established it for a classical ideal (i.e., non-interacting) gas. Dobrushin and Tirozzi [74] used in mid 70's the CLT to demonstrate the equivalence for lattice gas models. More general proofs of the the ensemble equivalence date to 90's and are connected with a concept closely related to the CLT, namely with a *large deviation principle* [68, 78, 97, 198].

We may note in passing that the independence of the distribution P_N or f_N (in the large-N limit) on actual form of the single-event distribution is one of the key hallmarks of the CLT. It is akin to the universality hypothesis of critical phenomena in which short-range details of a system do not affect its large-scale properties (e.g., order parameter) [41, 295]. As we shall see in Chapter 3.5, the same type of universality typifies random walks with

short-range memory and/or correlation. The only property that will be relevant is that the first two momenta are finite. All such random walks can be then universally described in the large-time limit by a *diffusion* process.

1.1.2 *Heavy-tailed distributions*

Despite rather mild assumptions which enter the CLT (cf. also Appendix B), there is a myriad of statistical and stochastic systems that violate strict Gaussian behavior regardless the fact that their macroscopic observables can be viewed as emerging from a large number of seemingly random effects. Let us look, for instance, on statistical systems depicted in Fig. 1.1.

Each of them is a result of a large collection of ostensibly random effects and yet the cumulative probability density function (PDF) exhibits polynomial (or heavy) tails. Such a non-Gaussianity should be attributed to a violation of one of the two basic premises of the CLT, that is, there either must exist non-trivial correlations among entering random variables (i.e., influences are not quasi-independent) or the ensuing first two momenta of individual distributions are not finite. Such violations can be observed in:

- *Systems with long-range correlations* — e.g., cellular automata or self-gravitating systems.
- *Systems with long-time memories* — e.g., financial markets or various chemical clocks
- *Non-ergodic systems* — e.g., glassy systems or (most) systems at low enough temperature.
- *Systems with non-Gaussian domains of attraction* — e.g., systems with Lévy noise or entangled Quantum Mechanical systems

Many systems, which violate the CLT clearly belong to more than one category. For instance, systems with long-time correlations or non-ergodic systems exhibit frequently also long-range correlations.

In the following subsection I will answer the question concerning the last premise of the CLT. In particular, I will show how the CLT should be generalized in order to account for random variable with PDF without first two momenta.

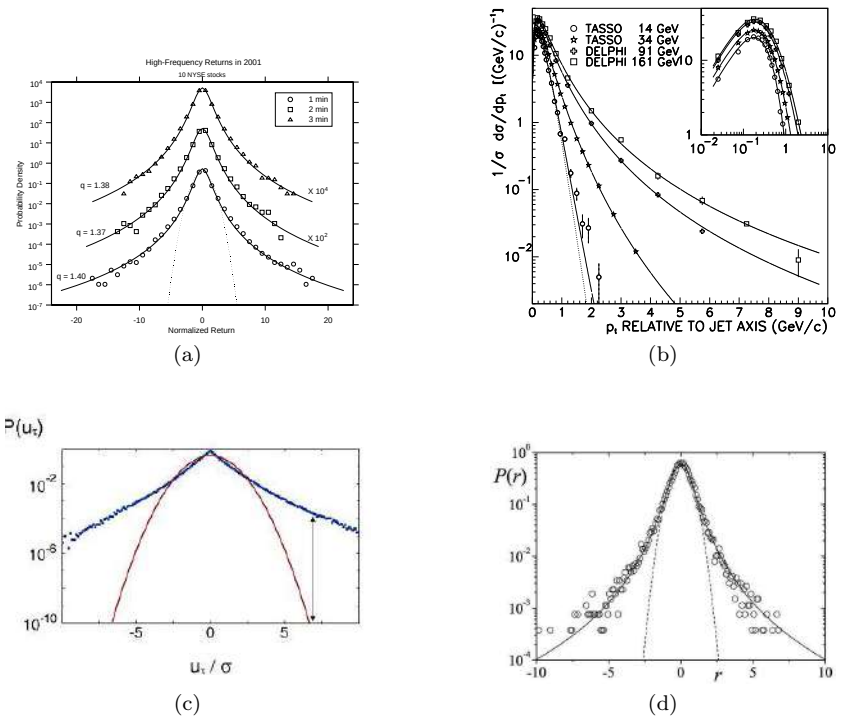


Fig. 1.1: Some examples of heavy-tailed distributions in real systems: (a) Empirical PDF's (points) and q-Gaussians (solid lines) for normalized returns of the 10 top-volume stocks in the NYSE in 2001. The dotted line is the Gaussian distribution. The 2- and 3-min curves are moved vertically. See Ref. [316]. (b) Distributions of transverse momenta p_t of hadronic jets produced in electron-positron annihilation. The dotted line is the Gaussian distribution. See Ref. [313]. (c) PDF of changes of wind speed for $\tau = 4$sec is represented by the squares. The solid line is the Gaussian distribution with the same standard deviation, see Ref. [42]. (d) PDF of returns, $P(r)$, versus return, r. The symbols represent $P(r)$ for the Dow Jones Industrial daily return index from 1900 until 2003. The solid line represents the best q-Gaussian numerical fit. See Ref. [317] for further details.

1.1.3 *Generalized central limit theorem*

In Section 1.1.1 we have seen that the Gaussian distribution is an attractor (or better fixed point in the functional space of PDF's) under addition of independent identically distributed random variables. In Appendix B I have, in addition, stated a number of useful and interesting generalizations of the CLT, which take into account various degrees of correlation among

entering random variables, however, all of these CLT's assume, in one way or another, assume that (at least) first two moments are finite.

Nevertheless, as we have observed in the previous sub-section, a number of real-world stochastic processes violate strict Gaussian behavior despite the fact that they emerge from a large number of seemingly random influences. As already mentioned, such a non-Gaussianity must be attributed to a violation of one of the premises of the CLT, i.e., either to non-trivial correlations among entering random variables (i.e., influences are not (quasi-)independent) or to non-existence of first two momenta of individual random variables. I shall concentrate here on the violation of the second premise.[1] In this connection a question arises, whether there exist a variant of the CLT that can account for an observable appearance of statistical systems with underlying heavy-tail distributions. In particular, one would like to address the entire class of the underlying single-event distributions that do not have finite second and/or first momenta, i.e., in 1D they posses long, inverse-power-law tails:

$$p(x) \;\sim\; \frac{1}{|x|^{1+\alpha}}, \qquad 0<\alpha<2 \qquad (|x| \;\to\; \infty)\,. \tag{1.7}$$

Such a *generalized* CLT was historically formulated independently in 1950's by B.V. Gnedenko [105, 106] and P. Lévy [196].

Before I embark on the actual formulation of the generalized CLT, I will mention two closely related important concepts, that will be needed in the following considerations. These are: 1) the notion of *stable distributions* and 2) Lévy distributions. Roughly speaking, stable distributions represent a class of probability distributions that do not change their functional form under convolution. In other words, for such distributions the statistical law of sum

$$\begin{aligned} f_N(x) \;&=\; \int p(x_1)p(x_2)\cdots p(x_N)\,\delta\left(\sum_{i=1}^{N} x_i - x\right)\prod_{i=1}^{N} dx_i \\ &=\; [p*p*\ldots*p](x) \;=\; [p^{*N}](x)\,, \end{aligned} \tag{1.8}$$

has exactly the *same shape* as the elementary single-event distribution $p(x)$. The fact that two distributions have the "same shape" means that one can find a (generally N-dependent) translation and dilatation of x such that the two probability laws coincide, i.e.

$$f_N(x)dx \;=\; p(y)dy \qquad \text{with} \quad y \;=\; a_N x + b_N\,. \tag{1.9}$$

[1] A violation of the first premise will be briefly discussed in Section 1.2.7

The most general class of stable distributions is the family of the so-called *Paretian* or *Lévy (stable) distributions*. Due to the lack of a closed form for PDF's for all but three distributions (see further), the stable distributions are most easily defined via their *characteristic function* — the inverse Fourier transform of the PDF. The latter have the generic form given by the following *Lévy* or *first Lévy–Khintchine theorem* [45, 106, 171, 195]:

Theorem 1.2. *A probability density $L_{\alpha,\beta}(x)$ is stable iff* [2] *the logarithm of its characteristic function*

$$\tilde{L}_{\alpha,\beta}(\xi) \;=\; \int_{-\infty}^{\infty} dx\, L_{\alpha,\beta}(x)\, e^{i\xi x}\,,$$

has the form

$$\ln \tilde{L}_{\alpha,\beta}(\xi) \;=\; i\gamma\xi - c|\xi|^{\alpha}\,(1 + i\beta\,\mathrm{sgn}(\xi)\,\omega(\xi,\alpha))\;,$$

where the parameters γ, c, α and β are real and take the values:

- *γ is arbitrary*,
- $c \geq 0$
- $\alpha \in (0, 2]$,
- $\beta \in [-1, 1]$,

and the function $\omega(\xi, \alpha)$ is given by

$$\omega(\xi,\alpha) \;=\; \begin{cases} -\tan(\pi\alpha/2) \text{ for } \alpha \neq 1\,, \\ (2/\pi)\ln|\xi| \quad \text{for } \alpha = 1\,. \end{cases}$$

□

In mathematical physics, the logarithm of a characteristic function is known under the name a *cumulant generating function*. The proof of the Lévy theorem (possibly formulated differently) can be found in numerous books on probability theory. Due to its quite complex nature I can point the interested reader, e.g., to Ref. [84] for more details.

Note that stable distributions require four parameters for complete description. Their meaning can be easily understood. First, for large $|x|$ the Lévy distributions fall off with the characteristic asymptotic power-law (see Appendix B.4)

$$L_{\alpha,\beta}(x) \;\sim\; \frac{1}{|x|^{1+\alpha}}, \qquad |x| \to \infty\,, \tag{1.10}$$

[2] Generally, the conjunction "iff" refers to "if and only if", i.e., to a necessary and sufficient condition.

The parameter α thus gives the tail exponent also called the tail index or index of stability. When $\alpha < 2$, the variance is infinite and the tails are asymptotically fat (or heavy), i.e., they exhibit a power-law behavior. The mean is finite provided that $\alpha \in (1, 2]$. All higher moments clearly diverge for $\alpha \in (0, 2]$. The case $\alpha > 2$ has no mathematical sense since the Fourier transform gives only semidefinite functions which are not suitable to represent PDF's [45]. The parameter γ gives the peak position, and specifically for $\alpha \in (1, 2]$ it coincides with the mean value[3]. The parameter β is called a skewness parameter and it describes how much the distribution is asymmetric under the parity transformation $x \leftrightarrow -x$ at $\gamma = 0$. Finally, the parameter c is a scale factor which characterizes the width of the distribution.

It is worth noting that a simple replacement of $x - \gamma$ with $c^{1/\alpha}x$ gives (provided $\alpha \neq 1$) the stable distribution that is no longer dependent on γ and which is multiplied with a simple normalization constant $c^{-1/\alpha}$, indeed

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi\, \exp[-i(x-\gamma)\xi - c|\xi|^{\alpha}\,(1 - i\beta\,\mathrm{sgn}(\xi)\omega(\xi,\alpha))]$$
$$\overset{x-\gamma\to c^{1/\alpha}x}{=} \frac{1}{2\pi c^{1/\alpha}}\int_{-\infty}^{\infty} d\xi'\, \exp[-ix\xi' - |\xi'|^{\alpha}\,(1 - i\beta\,\mathrm{sgn}(\xi')\omega(\xi',\alpha))]. \quad (1.11)$$

Here $\xi' = c^{1/\alpha}\xi$. We thus see that the parameters γ and c are only responsible for shifting the origin and rescaling the abscissa but they do not alter the shape. Parameters α and β are thus the only ones that are responsible for the shape and the other properties of stable distributions. For $\alpha = 1$ the replacement $x - (\gamma + 2\beta c \ln c/\pi)$ with cx does the same job.

For these reasons one typically uses only two parameters, to specify stable distributions. The members of the two-parametric class of PDF's $L_{\alpha,\beta}(x)$ are known as Paretian or Lévy (stable) PDF's. Sometimes also the name *α-stable distributions* is used. Among the most prominent examples of Lévy stable distributions belong:

Gaussian distribution: By setting $\alpha = 2$, $\beta \in [-1, 1]$, $c = \sigma^2/2$ and $\gamma = m$, I obtain the characteristic function

$$\tilde{L}_{2,0}(\xi) \;=\; \exp\left(im\xi - \sigma^2\xi^2/2\right), \quad (1.12)$$

which after the Fourier transform gives the Gaussian distribution

$$p_G(x) \;\equiv\; L_{2,0}(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}\,. \quad (1.13)$$

[3]This follows directly from the definition of the characteristic function.

So, we see hat the Lévy stable distributions exhibit a crossover from a power-low decay (with $\alpha < 2$) to the exponential tail (for $\alpha = 2$).

Lorentz–Cauchy distribution: By setting $\alpha = 1$, $\beta = 0$, $c \geq 0$ and $\gamma = m$, I obtain the characteristic function

$$\tilde{L}_{1,0}(\xi) \;=\; \exp\left(im\xi - c|\xi|\right), \tag{1.14}$$

which after the Fourier transform gives the Lorentz–Cauchy distribution

$$p_{LC}(x) \;\equiv\; L_{1,0}(x) \;=\; \frac{c}{\pi\left(c^2 + (x-m)^2\right)}, \quad x \in \mathbb{R}. \tag{1.15}$$

Lévy–Smirnov distribution: when I set $\alpha = 1/2$, $\beta = 1$, $c \geq 0$ and $\gamma = m$, I obtain the characteristic function

$$\begin{aligned}\tilde{L}_{1/2,1}(\xi) \;&=\; \exp\left(im\xi - \sqrt{c|\xi|}(1 - i\,\mathrm{sgn}(\xi))\right) \\ &=\; \exp\left(im\xi - \sqrt{-i2c\xi}\right),\end{aligned} \tag{1.16}$$

which after the Fourier transform gives the Lévy distribution

$$p_{LS}(x) \;\equiv\; L_{1/2,1}(x) \;=\; \sqrt{\frac{c}{2\pi}}\,\frac{\exp\left(-\frac{c}{2}(x-m)^{-1}\right)}{(x-m)^{3/2}}, \quad x \geq m. \tag{1.17}$$

These three PDF's are also the only ones that can be written in a closed form in terms of elementary functions, though there are also special situations where a closed form of the PDF is reachable in terms of special functions [84, 239].

Now we are in position to proceed to the generalized CLT. One of the frequently used formulation is the following [45]:

Theorem 1.3. *Let $\{X_k\}$ be a sequence of mutually independent random variables with a common distribution. Consider the stochastic process $S_N = X_1 + X_2 + \cdots + X_N$. Suppose further that the PDF of the random variable X_1 has an asymptotic tail given as*

$$p(X_1 = x) \;\sim\; \begin{cases} C_-|x|^{-(1+\alpha)}, \text{ as } x \to -\infty, \\ C_+|x|^{-(1+\alpha)}, \text{ as } x \to +\infty. \end{cases}$$

Define the parameter

$$\beta \;=\; \frac{C_+ - C_-}{C_+ + C_-}.$$

and scaled random variable $\hat{S}_N = S_N/N^{1/\alpha}$. *Then* $\hat{S}_N$ *has a density function* $f_N(x)$ *which satisfies the relation*

$$\lim_{N\to\infty} f_N(\hat{S}_N = x) \;=\; L_{\alpha,\beta}(x)\,,$$

i.e., $f_N(x)$ *belongs to the attraction basin of the Lévy stable distribution of the tail exponent* α *and skewness parameter* β.

A two comments concerning the generalized CLT are in order. The Theorem 1.3 states that the (normalized) sum of heavy-tailed distributions with the tail exponent α has as the limit distribution — the Lévy stable distribution, with the same tail exponent. The theorem is not applicable if the involved distributions have different tail exponents at spatial infinities $x \to \pm\infty$. Similarly, the theorem implies that a sum of two independent random variables with α-stable distributions is again stable distribution with the same tail exponent α. This invariance property, however does not hold if the two α's are different.

The generalized central limit theorem started to enter world of physics rather slowly. The first attempt (I am aware of) came from B. Mandelbrot[4], who in 60's investigated statistical systems with no finite second momenta in the context future-price forecast in financial models [212, 213, 215]. It should be stressed that distributions without second and/or first momenta were used as a golden thread in much of Mandelbrot works (see, e.g., [216–219]). By Mandelbrot's own admission [220], the whole concept of fractal mathematics stemmed from his effort to correctly mathematically describe the observed scaling in financial prices.

The aforementioned works remained largely overlooked both by the physics and quantitative-finance community. With few exceptions, such as *anomalous diffusion* or *self-organized criticality* which both started to be experimentally and theoretically analyzed already in 70's and 80's (see, e.g., Refs. [12, 120, 121, 175]), the real boom in the practical employment of the generalized CLT came at the turn of the 80's to 90's. The catalyst has been massive infusion of ideas from generalized thermostatistics of Tsallis, from theory of complex dynamical systems, as well as from improvements in observational and numerical sensitivity that allowed to examine large data sequences.

Let me finally stress that there are clearly other heavy-tailed alternatives to Lévy stable distributions. These include, for instance, Student's

[4]Perhaps not so surprisingly, Benoit Mandelbrot was a PhD student of Paul Lévy at the École Polytechnique in Paris.

t-distributions, normal inverse Gaussian distributions, Tsallis-type distributions or hyperbolic distributions. But even though the Lévy stable distributions are not universal, the generalized CLT provides them with a good reason why they should be expected in nature more than any other heavy-tailed PDF's. Namely, the stable laws are the only possible limit distributions for (properly normalized) sums of independent, identically distributed random variables.

1.2 MaxEnt and generalized entropies

Maximal entropy principle (or MaxEnt for short) is presently a key paradigm in statistical physics and probability theory. In both these fields it is related to the issue of *inference* about the most "representative" distribution for a system at hand which is compatible with whatever constraints (or knowledge) one might control. Since its inception, both statistical physics and probability theory have approached the MaxEnt principle differently — with different underlying justifications and with different objectives in mind. In the following I will briefly discuss the both respective routes and their merger in what is currently known as Jaynes' MaxEnt principle. I will also point out that the MaxEnt provides, among others, a natural arena in which some of heavy-tailed distributions find their conceptual justification.

1.2.1 *Entropy in statistical physics and thermodynamics*

The entropy concept was originally introduced by Clausius [59, 131] in the thermodynamics. By analyzing a Carnot engine he was able to identify a new *state function* which never decreases in isolated systems. Clausius showed that entropy of a given thermodynamic state can be interpreted as the measure of useful energy (i.e., energy available for work) that cannot be extracted from that state. So the increase in the entropy reflects a loss of useful energy in the course of system's evolution.

Modern axiomatic underpinning for thermodynamic entropy was provided in 1908 by the German-born Greek mathematician, Constantin Carathéodory [52] and later elaborated by Chandrasekhar [57], Buchdahl [47], Landsberg [191] and others. In contrast to Clausius' approach, Carathéodory's axiom is not directly rooted in any experience or experiment but when used in conjunction with purely mathematical theorem on integrability of Pfaffian differential forms (which is also due to Carathéodory)

one can derive all of the consequences (including existence of entropy), which normally follow from the Claussius approach. Carathéodory's approach is, however, more convenient in proving the existence of entropy for thermodynamic systems with 3 or more state variables [190].

The microphysical origin of Clausius' phenomenological entropy was clarified more than 20 years later in works of Boltzman and (yet later) Gibbs. In particular, Gibbs associated Clausius entropy with the number of allowed microscopic states compatible with a given observed macrostate. The ensuing so-called *Boltzmann–Gibbs entropy* (BGE) has the form

$$S_{BG}(\mathcal{P}) \;=\; -k_B \sum_{x\,\in\,\mathfrak{X}}^{W} p(x)\,\ln p(x)\,, \tag{1.18}$$

where $k_B = 1.3806488(13)\times 10^{-23}\mathrm{J/K}$ denotes Boltzmann's constant, $\mathfrak{X}$ is the set of all accessible microstates compatible with whatever macroscopic observable (state variable) one controls and W denotes the number of such microstates. The explicit form (1.18) of BGE conceptually follows from Gibbs' choice of the so-called *coefficient of probability* $P_\Gamma(\Omega)$ of the Γ-space (i.e., the macrosystem phase space) considered [101]. Gibbs made the choice so that the quantity $P_\Gamma(\Omega)d\Omega$ (here $d\Omega$ is an infinitesimal volume in Γ-space) would represent the ratio of the number of the ensemble systems contained in $d\Omega$ around the point Ω in Γ-space. With this Gibbs gave the definition of the entropy as [101]

$$S_G(\Omega) \;=\; -k_B \int_\Gamma P_\Gamma(\Omega)\,\ln P_\Gamma(\Omega)\,d\Omega\,. \tag{1.19}$$

Gibbs was able to show that in cases when the correct equilibrium ensemble is used then S_G reduces to the usual thermodynamic entropy.

Gibbs's entropy (1.19) had its precursor in Boltzmann's H-theorem. In the framework of his kinetic theory of gasses Boltzmann was able to identify a new function, the so-called H-function (also Eta-function)

$$\mathsf{H}(t) \;=\; \int_\mu f(\mathbf{x},\mathbf{p},t)\ln f(\mathbf{x},\mathbf{p},t)\;d\mathbf{x}d\mathbf{p}\,, \tag{1.20}$$

which *statistically decreases* in time. Function $f(\mathbf{x},\mathbf{p},t)$ appearing in (1.20) is a *single-particle* distribution in μ-space (i.e., a single-particle phase space). With this Boltzmann defined the entropy as $S_B(t) = -k_B\mathsf{H}(t)$. Note that for a system of N statistically independent particles, H is related to the Gibbs entropy through identity

$$S_G \;=\; -Nk_B\mathsf{H}\,. \tag{1.21}$$

It should be stressed that the equivalence between Boltzmann–Gibbs and Clausius entropy is established only when the conditional extremum $\mathcal{P}_{\rm ex}$ of S_{BG} subject to the constraints imposed by observed *state variables* is inserted back into S_{BG}. As long as this MaxEnt prescription is utilized, S_{BG} turns out to be a thermodynamic state function and not mere functional on a probability space [101].

The reason why one should strive for MaxEnt in statistical physics is closely related to the ergodic theorem. The ergodic theory is presently an independent branch of mathematics with many applications in information theory, statistics (e.g., the Strong Law of Large Numbers), number theory (e.g., Gelfand's problem), differential geometry, functional analysis, etc., much of which seems, however, unrelated to contemporary physics. In statistical physics, the ergodic hypothesis was introduced by L. Boltzmann in his 1877 paper [44] to prove equipartition of energy in the kinetic theory of gases. Boltzmann's ergodic hypothesis is frequently stated as the condition that the average of any quantity over all the microstates of a system equals the time average of that quantity in any one example of the system (i.e., when different initial conditions are used). To facilitate the concept of probability, Boltzmann conjectured that all phase space cells on a given energy hypersurface are equally probable — this is epitomized in his famous *Stoßzahlansatz*. Unfortunately, with the rigor that would satisfy mathematics community, classical dynamics has not been able to prove that the path of a generic system in phase space would move through all the cells, let alone spend equal time in each cell. The difficulty stems from the complexity of the underlying molecular/atomar motion resulting from the chaotic nature of particle collisions. For a system to be ergodic, the particle motions, must in some sense, be sufficiently mixed up or random. Though in a strict mathematical sense chaoticity of interactions does not imply ergodicity, it is likely that chaos explains the practical success of statistical mechanics. To account for billions of collisions that happen per particle per second in a usual gas[5] one must make a gross simplifications hoping that they capture the essence of the actual atomic disordered behavior. To this end, such properties as mixing, equidistribution or K-flow, have been extensively studied (see, e.g., [253, 308]).

From statistical physics point of view, the most important example of the ergodic theorem is the Birkhoff (or Birkhoff–Khinchin) theorem [38]. Before I state it, let me set the requisite terminology and compile some

[5]Note that a typical mean free time in gases is ~ 0.1ns. For instance in Ar is the mean free time 0.165ns, in N_2 it is 0.130ns, while in CO_2 only 0.108ns.

basic facts. As a rule, ergodic theory studies dynamical systems in terms of two fundamental mathematical concepts: a *measure space* $(X, \mathcal{F}, \mu)$ and a *measure preserving (time) map* T: $X \to X$.[6] I discuss the measure space in Appendix B.5, and so I will not dwell into at this stage. Let us now assume that we have the σ-algebra $\mathcal{F}$ and measure μ. One may then define measure preserving map as follows: first, one defines that T is measurable if for any $\mathcal{A} \in \mathcal{F}$ follows that $T^{-1}\mathcal{A} = \{x : T(x) \in \mathcal{A}\} \in \mathcal{F}$. That is, T is measurable if and only if the preimage of every measurable set under T is again measurable. The map T is then said to be measure preserving with respect to μ (or equivalently μ is T-invariant), when $\mu(T^{-1}\mathcal{A}) = \mu(\mathcal{A})$ for all $\mathcal{A} \in \mathcal{F}$. (If the map T is invertible this is the same as saying that $\mu(\mathcal{A}) = \mu(T\mathcal{A})$.) Clearly, a function $f : X \to \mathbb{R}$ transforms under action of T as $f(x) \mapsto Tf(x) = f(T(x))$. Generally, for the k successive T-transformations acting on $f(x)$ one has $T^k f(x) = f(T^k(x))$. When k is extended to the real axis $\mathbb{R}$, then $T^k \mapsto T_t$, $t \in \mathbb{R}$ and the family $\{T_t\}$ is called a (continuous) measurable flow in X. The measurable flow is measure preserving w.r.t. μ if for any $\mathcal{A} \in \mathcal{F}$ follows that $\mu(T_t^{-1}\mathcal{A}) = \mu(\mathcal{A})$ for all t.

In statistical physics the role of X is typically taken over by the phase space Γ and the transformation T_t represents the time evolution by a time t. For instance, for systems where the time evolution is driven by the Hamiltonian $H(\mathbf{p}, \mathbf{q})$ with $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ and $\mathbf{q} = (q^1, q^2, \ldots, q^N)$ the time transformation on a phase-space function $f(\mathbf{p}, \mathbf{q})$ can be explicitly written as

$$T_t f(\mathbf{p}, \mathbf{q}) \;=\; e^{t\mathcal{L}} f(\mathbf{p}, \mathbf{q}) \;=\; f(e^{t\mathcal{L}}(\mathbf{p}, \mathbf{q}))\,, \tag{1.22}$$

where the linear operator

$$\mathcal{L} \;=\; \sum_{n=1}^{N} \left\{ \frac{\partial H}{\partial q^n} \frac{\partial}{\partial p_n} - \frac{\partial H}{\partial p_n} \frac{\partial}{\partial q^n} \right\}, \tag{1.23}$$

is the *Liouvillian* of the system. The measure μ can be in this case taken to be the Lebesgue measure on Γ. The Liouville theorem then ensures that the above Hamiltonian flow $\{T_t\}$ is *measure preserving* w.r.t. μ on Γ.

After this prelude I can state the Birkhoff ergodic theorem. One of its frequently used versions is formulated as follows [38, 168]:

Theorem 1.4. *Let X be the configuration space of a physical system with a σ-finite measure μ and a function f: $X \to \mathbb{R}$. Let further T: $X \to X$ be*

[6] In ergodic theory it is customary to call T as *endomorphism*. If T is, in addition, reversible then the name *automorphism* is habitually used.

measure preserving time transformation. For any $f \in L^1(X,\mu)$ *the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} f(T^i(x)) = \langle f \rangle_x \,.$$

(i.e., the time average along a trajectory) exists for almost all $x \in X$*. If, in addition, the measure* μ *is ergodic (i.e., for all* $f \in L^1(X,\mu)$ *is* $\langle f \rangle_x$ *constant for almost all* x*) then for almost all* $x \in X$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} f(T^i(x)) = \int_X f(x)\mu(dx) \,.$$

This means that

$$\mu\left(x \in X\colon \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} f(T^i(x)) \neq \int_X f(x)\mu(dx) \right) \;=\; 0 \,.$$

The version that is in its spirit closest to the original Boltzmann ergodic hypothesis is Birkhoff's version for a measurable flow $\{T_t\}$ in X. In this case is the Birkhoff theorem states that [168]

Theorem 1.5. *Let* X *be the configuration space of a physical system with a* σ*-finite measure* μ *and a function* $f\colon X \to \mathbb{R}$ *such that* $f \in L^1(X,\mu)$*. For a measurable flow* $\{T_t\}$ *in* X *the limit*

$$\lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{T+t_0} f(T_t(x))dt \;=\; \langle f \rangle_x \,,$$

exists for almost all $x \in X$ *and is* t_0 *independent. If, in addition, the measure* μ *is ergodic then for almost all* $x \in X$

$$\lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{T+t_0} f(T_t(x))dt = \int_X f(x)\mu(dx) \,.$$

This means that

$$\mu\left(x \in X\colon \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{T+t_0} f(T_t(x))dt \neq \int_X f(x)\mu(dx) \right) \;=\; 0 \,.$$

The latter is an extremely important property for statistical mechanics, since it implies that for a typical orbit of an ergodic dynamical system, *time averages* equal *space averages*. In fact, it is the time averaged (or coarse-grained) behavior that is experimentally observed and tested but it is the phase-space average that can be mathematically dealt with.

The proof of the ergodic theorems was originally given by Birkhoff in Ref. [38]. Simpler (but still rather technical) proof was presented by A. Kolmogorov which was reviewed in the book of Khinchin [168]. The proof of the above stated versions can be found, for instance, in Refs. [255, 291].

Few comments are now in order. It should be first stressed that the configuration space X in statistical physics does not need to be (and as a rule is not) the entire phase-space. Typically it is only *allowed* part of it. By the allowed part I mean the set of all points in the phase space, which satisfy macroscopic constraints, such as constancy of energy, total linear momentum, total angular momentum or other integrals of the motion.

Second, Birkhoff's theorem does not ensure the existence of T-invariant ergodic measure and, in fact, the existence of such a measure in not guaranteed for general X's. However, if X is a *compact metric* space (e.g., phase space of an isolated Hamiltonian system) then for any *continuous* map $T\colon X \to X$ there exists at least one T-invariant measure (this results follows from the Krylov–Bogolyubov theorem [43, 236]). In addition, in the set of all T-invariant measures exists at least one T-ergodic measure (this follows from the Krein–Milman theorem [185]). In the special case when there is only one T-invariant measure, there exists a unique T-ergodic measure (Furstenberg's theorem [94]).

Third, if the above ergodic relation holds, then $\langle A\rangle_x$ does not depend neither on the initial position x (excluding a set of measure zero) nor on the initial time t_0. In particular, Birkhoff's theorem asserts that the time average is the same for almost all initial points (except possibly for a set of measure zero): statistically speaking, the system that evolves for a long time "forgets" its initial state. Such a behavior can be expected, e.g., in the (macroscopic) equilibrium. The premise that $\langle A\rangle_x \equiv \langle A\rangle$ corresponds at the fixed value of energy to the equilibrium value of A (for each observable A) constitutes the so-called *microcanonical ensemble* assumption and the corresponding μ is known as the *microcanonical measure*.

Fourth, if I take $A = \chi_C$, where

$$\chi_C(x) \;=\; \begin{cases} 1, \text{ for } x \in C \in \mathcal{F}; \\ 0, \text{ otherwise,} \end{cases} \tag{1.24}$$

is the *indicator function* of some set $C \subseteq X$, then the time mean $\frac{1}{n}\sum_{i=0}^{n} \chi_C(T^i(x))$ represents the relative frequency of visits to C of the trajectory staring at the point x and evolving in the time interval $[0, n]$. By Birkhoff's ergodic theorem the limit value of such frequencies as $n \to \infty$ exists and, in addition, it equals to the measure $\mu(C)$. This is therefore

analogous to the strong law of large numbers in probability theory. In the case of a flow $\{T_t\}$ the time mean $\frac{1}{T}\int_0^T \chi_C(T_t(x))$ represents the fraction of time the particle spends in the set C during its evolution from initial time $t = 0$ to the final time $t = T$. In the large T limit this represents the relative time spent in the set C (i.e., the relative *sojourn* time in C). So the ergodic measure of a given set C is proportional to the sojourn time of the system on the set C.

Let me finally stress that the Boltzmann ergodic conjecture as originally stated by Boltzmann is false, since it holds only when the involved integral measure is T-invariant and ergodic. In fact, the Boltzmann ergodic conjecture is based on the implicit assumption that the typical timescale over which the observation (and hence the coarse-graining or averaging) is made is sufficiently large in comparison with the inner timescales of the microscopic processes, so that the system has sufficient time to explore all the allowed part of phase space. Despite much effort the existence of ergodic properties in Boltzman's sense has never been proven for most systems. By now is known that a large number of physical systems is certainly non-ergodic, e.g., system of anharmonic oscillators. Some other systems are technically ergodic in the large-time limit but in practice non-ergodic over typical physical time scales. The latter is typical in systems that exhibit slow aging — such as (spin) glasses [247], or represent spontaneous-symmetry broken systems in the ordered phase — such as ferromagnets bellow the Curie temperature [96, 247].

In statistical physics it is often tacitly assumed that systems in question are ergodic enough to justify the use of the principle of *equal weight* (or *equal a priori probability*) of allowed microstates. That is, in an isolated system in a thermal equilibrium each of microstates on a constant energy hyper-surface is realized with equal probability, i.e.

$$\begin{aligned} \mathfrak{f}(\mathbf{p},\mathbf{q}) \;&=\; \text{constant} \;=\; \lim_{\delta E\to 0}\left[\int_{E<H<E+\delta E} d\Omega\right]^{-1}, \\ \Rightarrow\; \mu(dx) \;&=\; \mathfrak{f}(\mathbf{p},\mathbf{q})\delta(H(\mathbf{p},\mathbf{q})-E)d\Omega\,. \end{aligned} \tag{1.25}$$

Above μ is know as the *microcanonical measure*. Because the average is taken over the energy hyper-surface one can with the help of (1.25) write

$$\begin{aligned} \langle\cdots\rangle \;&=\; \int_X \cdots \mu(dx) \;=\; \frac{1}{\mathrm{Vol}\,\Gamma}\int_\Gamma \cdots \delta(H(\mathbf{p},\mathbf{q})-E)d\Omega \\ &=\; \frac{1}{\mathrm{Vol}\,\Gamma}\int_\Sigma \cdots \frac{d\Sigma}{|\mathrm{grad}H|}\,, \end{aligned} \tag{1.26}$$

where Σ is the energy hyper-surface and $\mathrm{Vol}\,\Gamma$ is the *accessible* phase-space volume, and

$$|\mathrm{grad}H| \;=\; \left[\sum_n \left\{ \left(\frac{\partial H}{\partial q^n}\right)^2 + \left(\frac{\partial H}{\partial p_n}\right)^2 \right\}\right]^{1/2} . \tag{1.27}$$

Similar considerations are valid also for quantum-statistical systems [131, 187].

Though, in statistical physics the most fundamental ensemble is the micro-canonical ensemble, the actual calculations are notoriously difficult in this ensemble. For that reason it is typically more tractable to consider other (but equivalent) ensembles by allowing for a small fluctuations in system's energy or number of particles.[7] For instance, using the principle of equal weight and partitioning the energy between two systems (that exchange energy) — heat bath (reservoir) and system under consideration, one gets a canonical-ensemble distribution [173, 187]

$$\mu(dx) \;=\; \frac{1}{Z}\,\exp(-\beta H(\mathbf{p},\mathbf{q}))d\Omega\,. \tag{1.28}$$

where Z represents the *partition function* (also called *sum-over-states*; German *Zustandssumme*) and $\beta = 1/k_B T$. In cases when one must take into account the precise relation between a volume element in the Γ-space and the corresponding number of distinct states of the system then

$$d\Omega \;=\; \frac{d\mathbf{p}d\mathbf{q}}{C_N}\,. \tag{1.29}$$

Here the factor $C_N = h^{3N}$ for *distinguishable* particles and $C_N = N!h^{3N}$ for *indistinguishable* particles (N represents the number of particles constituting the system and $h = 6.62606957(29)10^{-34}$ Js is the (unreduced) Planck constant).

In the statistical physics the justification for the MaxEnt is related to the presumed validity of the ergodic hypothesis. In particular, natural tendency of dynamical systems to (almost)evenly distribute over available microstates naturally leads to an increasing trend in the value of the BGE in the course of time. The BGE reaches its maximum in thermodynamic equilibrium, i.e., when the microstates are "maximally" evenly distributed over microstates condition the observable constraints (cf. e.g. [188]). For

[7] For systems that are away from thermodynamic limit the equivalence between ensembles is broken and a proper justification for a given ensemble is not due to ergodic hypothesis but instead via information theory, see Sec. 1.2.3.

instance, the canonical ensemble distribution comes from the conditional extremum of

$$\delta S_{BG}(\mathcal{P}) \;=\; 0, \quad \sum_{n\,\in\,\mathfrak{X}}^{W} p_n E_n \;=\; \langle H\rangle \;=\; \text{const.}, \quad \sum_{n\,\in\,\mathfrak{X}} p_n \;=\; 1\,, \quad (1.30)$$

which uniquely implies the maximizer

$$p_n \;=\; Z^{-1}e^{-\beta E_n}, \quad Z \;=\; \sum_{n\,\in\,\mathfrak{X}}^{W} e^{-\beta E_n}\,. \quad (1.31)$$

By the same token, one can generate also other standard ensembles when other constraining conditions are employed. In this way one can generate, e.g., *grand-canonical*, *isobaric = pressure*, *rotational*, etc. ensembles [138–141, 187].

For the sake of completeness, I should mention that apart from the ergodic hypothesis one may also find justification for the MaxEnt by introducing another concepts from the chaotic dynamical systems such as the metric (or Kolmogorov–Sinai) entropy and topological entropy. Loosely speaking these entropies represent numerical indicators of the degree of complexity in the orbit structure of the evolving system in its configuration space. In certain cases they can be indeed identified with the BGE. Because this topic is far beyond the scope of this thesis, I will refrain from discussing it further. The interested reader can find some essentials, e.g., in Refs. [243, 283].

Let me close by observing that MaxEnt's in statistical physics and thermodynamics are of a different nature, albeit in thermodynamic equilibrium they typically lead to identical conclusions. On thermodynamics side the entropy reaches its maximum in an equilibrium state (provided system evolves in an adiabatically isolated fashion) not due to ergodic hypothesis but due to Clausius inequality (cf. e.g. [187]). From this point of view is the thermodynamic entropy less ambiguous in its scope of validity and mathematical rigor than the BGE. Usefulness of the BGE stems mostly from its applications beyond equilibrium. Apart from equilibrium thermodynamics is the BGE routinely used in near-to-equilibrium thermodynamics, for instance, in Green–Kubo relations [13, 188] or in Onsager's reciprocal relations [141, 188]. In the following two sections we will see that there is yet another, conceptually deeper justification for the MaxEnt which does not relies on the concepts borrowed from dynamics.

1.2.2 *Entropy in information theory*

Information theory concerns how efficiently one can encode information, compute, evade eavesdroppers, and communicate. One of the key goals of information theory is to construct an *optimal code*. By optimal code one means the shortest averaged code from which one can *uniquely decode* the source data. Optimality of coding was first solved by Shannon in his 1948 seminal paper [285, 286]. According to Shannon's *source coding theorem* [285, 286, 288], the quantity

$$H(\mathcal{P}) \;=\; -\sum_{x\,\in\,X} p(x)\,\log_2 p(x)\,, \tag{1.32}$$

corresponds to the averaged number of bits needed to optimally encode (or "zip") the source dataset X with the source probability distribution $\mathcal{P}(X)$ (more details will be presented in Chapter 2.1). On a quantitative level (1.32) represents (in bits) the minimal number of binary (yes/no) questions that brings us from our present state of knowledge about the system X to the one of certainty [8, 263, 285, 286]. It should be stressed that in Shannon's formulation X represents a discrete set (e.g., processes with discrete time), and this will be also the case here. Apart from the coding-theory based *operational* definitions, Eq. (1.32) has also several axiomatic underpinnings. Axiomatic approaches were advanced by Shannon himself in Refs. [285, 286, 288], Khinchin [169], Fadeev [82] an others [114]. The quantity (1.32) has became known as Shannon's entropy (SE).

1.2.3 *Jaynes' MaxEnt*

Jaynes and his so-called *subjectivist* school has followed a different line of reasoning in order to justify the MaxEnt. To them, the problem of statistical mechanics is basically a problem of statistical inference. As we have seen, Shannon's entropy is simply the measure of the uncertainty inherent in a preassigned probability scheme, and as such it has nothing to do with thermodynamic entropy *per se*, except that in the cases in which the probability distribution is known, or proven to be, "(grand-)canonical" one can identify the ensuing SE with the thermodynamic entropy of Clausius. Since the SE quantifies lack of knowledge (or our ignorance) about a system, the (grand-)canonical distributions naturally enter, according to Jaynes, the scene when one tries to infer the least biased distributions permitted by our knowledge of the mean energy (and mean number of particles). In this approach the ergodicity is not requisite. Jaynes' justification of MaxEnt

principle is sometimes also known as the *principle of maximum honesty*.

Already in mid 50's Jaynes pointed out that there is an intimate connection between Boltzmann–Gibbs entropy and Shannon's entropy [138, 139]. In fact, thermodynamics can be viewed as a specific application of Shannon's information theory: thermodynamic entropy may be interpreted (when rescaled to "bit" units) as the amount of Shannon information needed to define the detailed microscopic state of the system, that remains "uncommunicated" by a description that is solely in terms of thermodynamic state variables [46, 138, 139, 141, 302].

There have arisen a number of objections in connection with Jaynes' information-theoretic underpinning of thermodynamic entropy. These originated from the viewpoint that thermodynamics entropy is a measurable quantity (up to an additive constant), and as such it should be independent of observer's knowledge (different people have different amounts of ignorance). Any other definition of entropy that is observer-dependent is unacceptable.

In this connection I can quote Jaynes' answer to critique of Professor Uhlenbeck [141]: *The entropy of a thermodynamic system is a measure of the degree of ignorance of a person whose sole knowledge about its microstate consists of the values of the macroscopic quantities X_i which define its thermodynamic state. This is a completely 'objective' quantity, in the sense that it is a function only of the X_i, and does not depend on anybody's personality. There is no reason why it cannot be measured in the laboratory.*

Similar rebuttal to the criticism was given by Rothstein [273]: *Physical information and its associated entropy reduction, localized in the system to which the information refers, can be expressed via specification or constraints taken as a part of the description of the system ... Measuring a system and thus finding it to be in some state is formally equivalent ... to preparing it to be in that state, specifying it to be in that state, or constraining it in a manner that it can be in no other state (the state in question can, of course, be mixed).*

By now, it is well recognized [141, 138–140] that within the context of Shannonian information theory the laws of equilibrium statistical mechanics can be viewed as *inferences* based entirely on prior information that is given in terms of *expected values* of energy, energy and number of particles, energy and volume, energy and angular momentum, etc. This framework is known as *inductive inference* and its internal consistency and uniqueness was proved by Shore and Johnson in Ref. [290].

Since the SE differs only by a scaling factor from Boltzmann's and Gibbs

entropies one can obtain, for instance, the canonical ensemble distributions by looking for an extremum of the SE subject to $\langle H \rangle = \text{const.}$, which yields

$$S_G(\mathcal{P}) \;=\; k_B H(\mathcal{P}) \;=\; -k_B \sum_{n \in X} p_n \log p_n$$

$$\stackrel{\text{MaxEnt}}{\Rightarrow} \qquad p_n \;\propto\; e^{-\beta E_n} \,. \tag{1.33}$$

In the case that the SE is formulated of the μ-space, one arrives at

$$S_G(\mathcal{F}) \;=\; -N k_B \mathsf{H}(\mathcal{F}) \;=\; -N k_B \int_\mu d\mathbf{x} d\mathbf{p}\; f \log f$$

$$\stackrel{\text{MaxEnt}}{\Rightarrow} \qquad f(\mathbf{p}) \;\propto\; e^{-\beta \mathbf{p}^2/2m} \,. \tag{1.34}$$

which is the usual stationary solution of Boltzmann's H-function, i.e., Maxwell–Boltzmann distribution[8].

Let me finally note that, the close relation between information theory and the BGE was already anticipated by Szilard [302] and Brillouin [46] in their analysis of the *Maxwell demon*[9]. A modern connection of the information theory and BGE is phrased via the so-called *Landauer's principle* [189]. The latter basically states that when a single bit of information is erased (e.g., in a computer or in a black hole) then the amount of energy dissipated into environment is at least $k_B T \ln 2$, where T is the temperature of the erasing environment. Landauer's principle has been elaborated in the literature quite extensively. The debate culminated in works of Bennett [29, 30], Lloyd [202], Vedral [321] and others.

1.2.4 *Entropy and its basic properties*

Let me now briefly state some of the basic properties that are (due to identical mathematical structure) shared by both BGE and SE.

Among important properties of SE is its concavity in $\mathcal{P}$, i.e. for any pair of distributions $\mathcal{P}$ and $\mathcal{Q}$, and a real number $0 \leq \lambda \leq 1$ holds

$$H(\lambda \mathcal{P} + (1-\lambda)\mathcal{Q}) \;\geq\; \lambda H(\mathcal{P}) + (1-\lambda) H(\mathcal{Q}) \,. \tag{1.35}$$

[8] Entropies over density functions are not true entropies but rather entropy gains, cf. Ref. [144] and are normally called *differential entropies*. More on differential entropies will be said in Chapter 2.4.3.2.

[9] Maxwell demon, is a hypothetical being of intelligence but molecular order of size imagined to illustrate limitations of the second law of thermodynamics was invented by J.C. Maxwell

Eq. (1.35) follows from Jensen's inequality and a convexity of $x \log x$ for $x > 0$. Concavity is an important concept since it ensures that any maximizer found by the methods of the differential calculus yields an absolute maximum rather than a relative maximum or minimum or saddle point. At the same time it is just a sufficient (i.e., not necessary) condition guaranteying a unique maximizer. It is often customary to denote SE of the source[10] X as $H(X)$ rather than $H(\mathcal{P})$. Note that SE is generally not convex in X!

It should be stressed that the entropy (1.32) really represents a self-information: the information yielded by a random process about itself. A step further from a self-information offers the *joint entropy* of two random variables X and Y which is defined as

$$H(X,Y) \;=\; -\sum_{x\,\in\,X,\,y\,\in\,Y} p(x,y)\log_2 p(x,y)\,, \tag{1.36}$$

and which represents the amount of information gained by observing jointly two (generally dependent or correlated) statistical events.

A further concept that I will need is the *conditional entropy* of X given Y, which can be motivated as follows: Let us have two statistical events X and Y and let event Y has a sharp value y, then the gain of information obtained by observing X is

$$H(X|Y=y) \;=\; -\sum_{x\,\in\,X} p(x|y)\log_2 p(x|y)\,. \tag{1.37}$$

Here the conditional probability $p(x|y) = p(x,y)/p(y)$. For general random variable Y one defines the conditional entropy as the averaged Shannon entropy yielded by X under the assumption that the value of Y is known, i.e.

$$\begin{aligned} H(X|Y) \;&=\; \sum_{y\,\in\,Y} p(y) H(X|Y=y) \\ &= -\sum_{x\,\in\,X,\,y\,\in\,Y} p(x,y)\log_2 p(x|y)\,. \end{aligned} \tag{1.38}$$

From (1.38) it follows that

$$H(X,Y) \;=\; H(Y) \;+\; H(X|Y) \;=\; H(X) \;+\; H(Y|X)\,. \tag{1.39}$$

[10] In Shannon's theory *source* is typically a random variable, say X that has a discrete *alphabet* (a set of possible values or autcomes) $\mathcal{X} = x_1, \ldots, x_n$. Strictly speaking one should write Shannon's entropy as

$$H(\mathcal{P}_X) = H(X) = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)\,,$$

but I will stick to more conventional notation given by (1.32).

Identity (1.39) is known as additivity (or chain) rule for Shannon's entropy. In statistical thermodynamics this rule allows to explain, e.g., Gibbs paradox [318]. Because both $H(X|Y) \geq 0$ and $H(Y|X) \geq 0$, I get that

$$H(X,Y) \;\geq\; H(Y) \quad \text{and} \quad H(X,Y) \;\geq\; H(X)\,. \tag{1.40}$$

These inequalities (known as the *monotonicity* of SE) state that when more random variables are included then the entropy of the joint outcome gets never smaller. Or, equivalently, the entropy of the whole is at least as great as the entropy of the part.

Another relevant quantity that will be needed is the *mutual information* between X and Y. This is defined as:

$$I(X;Y) \;=\; \sum_{x\,\in\,X,\;y\,\in\,Y} p(x,y)\log_2 \frac{p(x,y)}{p(x)q(y)}\,, \tag{1.41}$$

and can be equivalently written as

$$I(X;Y) \;=\; H(X) \;-\; H(X|Y) \;=\; H(Y) \;-\; H(Y|X)\,. \tag{1.42}$$

This shows that the mutual information measures the average reduction in uncertainty (i.e., gain in information) about X resulting from observation of Y. Of course, the amount of information contained in X about itself is just the Shannon entropy:

$$I(X;X) \;=\; H(X)\,. \tag{1.43}$$

Notice also that from Eq. (1.41) follows $I(X;Y) = I(Y;X)$ and so X provides the same amount of information on Y as Y does on X. For this reasons the mutual information is not a useful measure to quantify a flow of information. In fact, the flow of information should be by its very definition directional.

It is also interesting to observe that since $I(X;Y) \geq 0$ (Gibbs's inequality, see Appendix D.2), I have that $H(Y|X) \leq H(Y)$ which can be perceived as a variant of the second law of thermodynamic. Namely, entropy increases when the influence of the boundary conditions gets lost in the course of time evolution.

In the following we will also find useful the concept of *conditional mutual entropy* between X and Y given Z which is defined as

$$\begin{aligned} I(X;Y|Z) \;&=\; H(X|Z) \;-\; H(X|Y,Z)\,, \\ &=\; I(X;Y,Z) \;-\; I(X;Y)\,. \end{aligned} \tag{1.44}$$

For further details on the basic concepts of Shannon's information theory, we refer the reader to classical books, e.g., Ash [8] and, more recently, Csiszár and Shields [64] or Cover and Thomas [62].

Let me note finally that strictly speaking Shannon's entropy can be written with an arbitrary multiplicative constant, i.e., as

$$H(\mathcal{P}) \;=\; -\kappa \sum_n p_n \log p_n \,.$$

If the multiplicative constant is chosen to be $\kappa = k_B$ then Shannon's entropy corresponds Gibbs entropy [101]. The presence of Boltzmann's constant k_B merely indicates the units in which Shannon's entropy coincides with Clausius thermodynamic entropy (should we choose p_n to be Gibbs's maximizer). Case with $\kappa = \log 2$ corresponds to Shannon's entropy measured in *bits*. Sometimes one chooses $\kappa = 1$ which represents Shannon's entropy measured in natural units — *nats*.

1.2.5 *World beyond Shannon*

As I have already mentioned, the SE provides via MaxEnt prescription *a least* biased PDF compatible with observable (or macroscopic) constraints. For instance, the usual thermodynamics ensemble PDF's are obtained without *ergodic* and *metric transitivity* assumptions, which are normally indispensable in ergodic theory.

Despite their heuristic appeal and conceptual simplicity, the arguments from Section 1.2 concerning the MaxEnt inference procedure have a loophole. The loophole resides in the fact that *information* in physics is a *richer* concept than Shannon's information measure can grasp. This fact was also recognized and stressed by Shannon himself: *It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field. ...* [287]

In following sub-sections I will illustrate this point with some selected information concepts that do not fall under the umbrella of Shannon's information theory, but first it will be convenient to consider Shannon's information measure from a slightly different angle of view.

1.2.5.1 *Shannon's information as syntactic information:*

Word "syntax" refers to the rules used for constructing, or transforming the symbols and words in a language (be it formal, programming or natural language). Particularly in linguistics the syntactic rules prescribe how words are combined to larger units — to phrases and sentences, that are well-formed (but not necessary meaningful) strings in a given language. So, for instance, the sentences: "portrait Rembrandt painted that a ..." and

"A portrait that Rembrandt painted ..." or "1+1 = 2" and "1 = 2+1" are from a syntactic point of view identical. The ensuing *syntactic information* then quantifies the amount of uncertainty involved in transmission and processing of well-formed syntactic strings.

In syntactic systems one can often affiliate with the "alphabet" (i.e. a set of building language units — usually words and hyphens) a probability distribution. Since the syntactic strings do not need to convey any meaning, one can consider the "alphabet letters" appearing in the string as statistically independent events. In this way one can endow the string with a statistical description through the product of probabilities associated with "alphabet letters" involved. For aforesaid syntactic systems Shannon's entropy is the pertinent quantifier of the syntactic information. Indeed, Shannon's entropy takes into account only the probability of observing specific "alphabet letters", so the information it encapsulates is information about the underlying probability distribution of the "alphabet", not about the meaning of the events themselves. Reverse is also true, Shannon's information theory, by its very formulation, deals only with information systems that are syntactic in their nature.

1.2.5.2 *Semantic information:*

Above mentioned syntactic information should be contrasted with the so-called *semantic information*. The latter is concerned with the "meaning" conveyed by a transmitted message or signal. In fact, the word "semantics" refers to the field of study that deals with meanings in languages. Here, by "language" I mean generically both formal, programming and natural language. Every language has apart from its own *syntax* also *semantics*. Roughly speaking, syntactic information is how to say something in a "grammatically correct" way, whereas semantic information has some sort of context or meaning behind what is said. The concept of semantic information was introduced by Bar-Hillel–Carnap [15], and Minsky [228, 229].

The difference between syntactic and semantic information can be illustrated, for instance, by the difference between DNA molecule and snowflake. In particular, DNA molecule contains semantic information because it has intrinsic meaning. Indeed, DNA stores the instructions needed to construct and operate organism. By putting DNA in right molecular environment — in the right semantic context, life happens! DNA has clearly also syntactic information epitomized by Shannon's entropy. Such information *per se* does not carry any meaning since it makes no distinction between mean-

ingful DNA sequences that encodes life, and random DNA sequences of equal length and equal probability distribution of nucleotides. By contrast, snowflakes contain only syntactic information which is encoded in the specific arrangements of their hexagonal shapes. These patterns have no semantic content, there is no meaning for anything beyond the structure itself. Snowflakes certainly do not code for, or symbolize anything, whereas genes (i.e., regions od DNA) most definitely do.

1.2.5.3 *Algorithmic information:*

As already mentioned, Shannon's entropy quantifies the syntactic information. It clearly ignores the meaning of the message/signal as the only relevant quantity is the probability associated with "alphabet letters". Question thus stands whether one can define some measure of semantic information. While the idea of semantic information is intuitively clear, it is difficult to implement it quantitatively. The most prominent exception is the concept of *algorithmic information* (or *complexity*) introduced in their seminal works by Chaitin [56] and Kolmogorov [180–182].

In its essence, the algorithmic information of a message/signal (typically binary) string is equivalent to the length of the shortest possible self-contained representation of that string. A self-contained representation is essentially a program — in some fixed but otherwise irrelevant universal programming language — that, when run, generates the original string. When the self-contained representation is shorter than the signal/message string one says that the string is *algorithmically compressible*. For instance a binary string

$$10101010101010101010101010101010101010101010\,,$$

can be represented by the simple instruction: print 10 twenty-two times. So, the (algorithmic) information content is very low because the sequence can be described by a very simple procedure or computer algorithm.

Often, a seemingly random (patternless) sequence might have a very low information content. For instance, the binary representation of π is

$$11001001000011111101101010100010001000010110\ldots\,,$$

which can clearly be generated via simple algorithm.

According to Chaitin [56], a sequence is random if it cannot be algorithmically compressed, i.e., the shortest description of a random sequence is simply the sequence itself. So, clearly π is not algorithmically random.

Clearly, disadvantage of the algorithmic information paradigm is that one can not (in most situations) decide about string compressibility. For instance, a portion of genom of the virus MS2 has the binary representation [65]

$$\ldots 0100011101110100100111001101011010111 0111 \ldots \ .$$

Here binary assignment of the base nucleotides is: $A = 00$, $U = 11$, $G = 01$ and $C = 10$. Is there a code within the genetic code? Nobody knows, though it is believed that "higher" organisms should have genomes with higher algorithmic entropy in order to store genetic information required for their biological function.

1.2.5.4 *Quantum information:*

Quantum theory brings yet another twist into the concept of information. In order to deal with information theoretic issues in quantum theory one needs to treat probability distribution over quantum states. This is embodied in the *density matrix* conventionally denoted as $\hat{\rho}$. For an n-dimensional Hilbert space (quantum state space) $\mathcal{H}$, a density matrix is $n \times n$ Hermitian trace-one positive semi-definite matrix. A rank one density matrix corresponds to the pure state $|a\rangle$ in which case $\hat{\rho} = |a\rangle\langle a|$ (which is nothing but the projector onto the subspace spanned by $|a\rangle$).

Density matrices emerge from quantum states in two qualitatively distinct ways. In the first case, the density matrices are associated with "classical" probability distributions over quantum states. That is, if a given system is prepared (by nature or by experimentalist) in a pure state $|a_i\rangle$ with probability p_i, then

$$\hat{\rho} \ = \ \sum_k p_k |a_k\rangle\langle a_k| \, . \tag{1.45}$$

An example might be provided by a system in thermal equilibrium where the probability of the system being in the microstate $|E_i\rangle$ with an energy E_i is proportional to the Boltzmann factor $e^{-\beta E_i}$.

In the second case, density matrices emerge through ignoring part of a pure entangled quantum state. Let me recall that two (or more) systems that are entangled have a definite quantum state (i.e., a pure state) when considered jointly, but to each of the systems separately one cannot assign a pure state. To illustrate this point, let me consider a bi-partite pure state defined on a tensor product system $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\dim \mathcal{H}_A = \dim \mathcal{H}_B = n$. In this case it is a matter of taste if I employ density-matrix or pure-state

description. By using the density matrix I can write

$$\hat{\rho}_{AB} \;=\; \left(\sum_{i,j} c_{ij}|a_i\rangle \otimes |b_j\rangle\right)\left(\sum_{p,q} \langle a_p| \otimes \langle b_q|c^*_{pq}\right). \tag{1.46}$$

If one can see (or control) only part A of the system, this part is described be the reduced density matrix

$$\hat{\rho}_A \;=\; \mathrm{Tr}_B(\hat{\rho}_{AB}) \;=\; \sum_{i,j} a_{ij}|a_i\rangle\langle a_j|\,, \tag{1.47}$$

with $a_{ij} = \sum_k c_{ik}c^*_{jk}$ (so a_{ij} is the Hermitian trace-one matrix with non-negative eigenvalues — as it should be). With the help of the Cauchy–Schwarz inequality I can write

$$\mathrm{Tr}_A(\hat{\rho}^2_A) \;=\; \sum_{ij} a_{ij}a_{ji} \;=\; \sum_{ij} |a_{ij}|^2 \;\leq\; (\sum_i a_{ii})^2 \;=\; 1\,, \tag{1.48}$$

so, unless all but one eigenvalues of a_{ij} are zero (i.e., unless $\hat{\rho}_A$ describes a pure state), the above inequality is not saturated. As a result, the quantum state $\hat{\rho}_A$ is generally not a pure state and the description via density matrix is compulsory. Similar argument applies for $\hat{\rho}_B$.

The quantity $\mathcal{P} = \mathrm{Tr}(\hat{\rho}^2)$ is called the *purity* of a state. A state is pure when its purity equals 1, and mixed otherwise. By employing Jensen's inequality for convex functions [118], I can estimate $\mathcal{P}_A$ from below, namely

$$\mathcal{P}_A \;=\; \sum_{ij} |a_{ij}|^2 \;=\; \sum_i \lambda_i^2 \;\geq\; \frac{1}{n}(\sum_i \lambda_i)^2 \;=\; \frac{1}{n}\,. \tag{1.49}$$

Here $\{\lambda_i\}_{i=1}^n$ is the spectrum of a_{ij}. A key point about Jensen's inequality is that it provides a clear criterion under which the inequality saturates. In particular, Eq. (1.49) is saturated if and only if all λ_i are equal, in which case $\lambda_i = 1/n$ for all i. This is the same as saying that $\hat{\rho}_A$ is a $1/n$ multiple of the identity matrix. The state $\hat{\rho}_A$ that saturates Eq. (1.49) is known as a *maximally mixed state* because it is a mixture where all states occur with the same probability. In addition, the latter is true in any orthonormal basis in which system A is measured.

A typical example of bi-partite entangled pure states that produce maximally entangled states is provided by the four Bell states [238]:

$$\begin{aligned} |\Phi^{\pm}\rangle \;&=\; \frac{1}{\sqrt{2}}\left(|0\rangle_A \otimes |0\rangle_B \;\pm\; |1\rangle_A \otimes |1\rangle_B\right), \\ |\Psi^{\pm}\rangle \;&=\; \frac{1}{\sqrt{2}}\left(|0\rangle_A \otimes |1\rangle_B \;\pm\; |1\rangle_A \otimes |0\rangle_B\right). \end{aligned} \tag{1.50}$$

In all four cases the reduced density matrix $\hat{\rho}_A = \frac{1}{2}(|0\rangle_A\langle 0|_A + |1\rangle_A\langle 1|_A) = \frac{1}{2}\hat{\mathbb{1}}$ and ensuing $\mathcal{P}_A = 1/2$. Similarly for $\hat{\rho}_B$ and $\mathcal{P}_B$. Note that both $\mathcal{P}_A$ and $\mathcal{P}_B$ saturate the lower bound and so the above Bell states are indeed maximally mixed.

Another alternative criterion used for quantifying pure states versus mixed states is based on Shannon's entropy (in the form of von Neumann's entropy), i.e., $H = -\mathrm{Tr}\,(\hat{\rho}\log_2\hat{\rho})$. In contrast to purity, Shannon's entropy is *minimal* for a pure state (actually zero[11]) and *maximal* for maximally mixed state (in which case it is $\log_2 n$ — for n-dimensional Hilbert space). One can easily check that for the Bell states (1.50) one has $H(\hat{\rho}_A) \equiv H(A) = H(\hat{\rho}_B) \equiv H(B) = 1$.

Information theory allows to establish connections between the concept of information and purity. To do so, one must, however, go beyond Shannonian paradigm. For instance, purity is trivially related to the so-called Tsallis–Havrda–Charvát entropy of the order two; $\mathcal{S}_2$ (cf. Chapter 1.2.7) and with Rényi information measure of the order two; $\mathcal{I}_2$ (cf. Chapter 1.2.8). Namely,

$$\mathcal{S}_2 \;=\; 1 \;-\; \mathcal{P}\,, \quad \mathcal{I}_2 \;=\; -\log_2\mathcal{P}\,. \tag{1.51}$$

One often employs also other two quantities, namely *tangle*

$$\tau(\hat{\rho}) \;=\; 2\mathcal{S}_2(\hat{\rho}_A) \;=\; 2\left(1 - 2^{-\mathcal{I}_2(\hat{\rho}_A)}\right), \tag{1.52}$$

and *concurence* $C = \sqrt{\tau}$. By expressing the criterion for mixed states in terms of information measures (i.e., entropies) rather than purity, one can regard mixing as a degree of uncertainty (or ignorance) about the pure states involved. Loosely speaking, the higher mixing entropy, the higher uncertainty about the weights that the respective pure states carry in the mixed state. Advantage of this information theoretic setting is that one can affiliate with mixed states an operational meaning furnished by various (Shannonian and non-Shannonian) coding theorems, see, e.g. [50, 51].

Apart from mixed-state diagnostic, generalized entropic measures can also be used as quantifiers (or degrees) of entanglement. I have pointed out that when a given pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is not entangled (i.e., is *separable*), then both reduced density matrices $\hat{\rho}_A$ and $\hat{\rho}_B$ correspond to

[11]It is often said that since in quantum theory the vacuum state is a pure state, the 3rd law of thermodynamics does not need to be considered as an independent postulate. This is true, provided certain assumptions are satisfied. Those assumptions include: a) statistical picture is considered as being more important than thermodynamical one, b) the Gibbs/Shannon/von Neumann's entropy functional is the correct measure of ignorance about an actual microstate.

pure states, i.e., $\mathcal{P}_A = \mathcal{P}_B = 1$. In the opposite case the pure state $|\psi\rangle$ is entangled. This naturally begs for the question: to what extend is a given state $|\psi\rangle$ entangled? There does not exist a unique answer to this question. Frequently used criterion (mainly for the bipartite systems) is the so-called *entanglement entropy*, which is nothing but the von Neumann entropy applied to the reduced density matrix $\hat{\rho}_A$ or $\hat{\rho}_B$.[12] When this is equal to zero one has a separable pure state, when it is $\log_2 n$ (i.e., the reduced density matrices are maximally mixed states), then $|\psi\rangle$ is said to be *maximally entangled*.

In general, the issue of separability (not to mention classification) of multipartite states is still unsolved problem. What is, however, well recognized is that entropies (both Shannon and non-Shannon-type) play a pertinent rôle in this task. Many of the axiomatic entanglement measures exist which are based on the concept entropy, such as relative entropy of entanglement [323], reversed relative entropy of entanglement [323] or squashed entanglement [324]. For instance, Rényi information measure of the order ∞ (cf. Chapter 1.2.8) has the form $\mathcal{I}_\infty(\hat{\rho}) = -\log_2 \lambda_{max}$. Here λ_{max} is the maximal value of the Schmidt vector, i.e. the component vector in the Schmidt decomposition of $\hat{\rho}$. If the Schmidt vector is ordered decreasingly and $\lambda_1 = \lambda_{max}$, then the state $|0\rangle \otimes |0\rangle$ is the separable pure state closest to $|\psi\rangle$ in the sense of the Fubini–Study distance. There the distance of $|\psi\rangle$ to the set of separable pure states is

$$D_{FS}^{\min} \;=\; \arccos(\sqrt{\lambda_{\max}}) \;=\; \arccos\left(2^{-\frac{1}{2}\mathcal{I}_\infty(\hat{\rho})}\right). \tag{1.53}$$

Important is also the so-called *Bures distance* which measures the distance from an analyzed state $\hat{\rho}$ to the set of separable states. This can be phrased (at least in some cases) in terms of the Rényi entropy [322].

Let me stress that quantum SE (von Neumann entropy) has a number of properties than are different from its classical counterpart. For instance, if A and B are two random variables then classically we have (see Eq. (1.40)) that $H(A, B) \geq H(A)$ and $H(A, B) \geq H(B)$. If I now consider any of the four Bell's states from (1.50) then $H(A) = H(B) = 1$. On the other hand, the joint system AB is a pure state and so $H(A, B) = 0$. Consequently, because of the entanglement the quantum entropy of the whole can be less

[12] In fact, due to Schmidt decomposition [77] it is immaterial whether one uses reduced density $\hat{\rho}_A$ or $\hat{\rho}_B$, both will provide the same entanglement entropy. Let me remind that the Schmidt decomposition ensures that that $\hat{\rho}_A$ or $\hat{\rho}_B$ have the same non-zero eigenvalues with the same multiplicities. The multiplicity of the zero eigenvalue (if present) may or may not be different. The latter is the case only when $\dim \mathcal{H}_A \neq \dim \mathcal{H}_B$.

than the quantum entropy of the part. Another (closely related) problem is related to the definition of a conditional quantum entropy. By using the classical definition of it, i.e. $H(A|B) = H(A,B) - H(B)$, we see that for entangled states this can be negative and so it cannot be understood as an average over (positive) entropies as in (1.38).

Let me finally mention that apart from issues related to entanglement, quantum mechanics adds also further layer of problems to the concept of entropy. For instance, in the presence of a conservation law the obtainable accuracy of the measurement of a physical observable is limited whenever the operator representing the observable does not commute with the conserved quantity. Any related consistent quantum-mechanical entropy should quantify our knowledge of such a difficult-to-measure observable with respect to a conserved quantity. In particular, it should vanish when the observable commutes with the conserved quantity (our ignorance can be completely removed). In this context a number of information measures were proposed. Among these, perhaps the most known candidate measure is the so-called Wigner–Yanase skew information [332].

1.2.6 *From Boltzmann–Gibbs statistics to heavy tails*

In the search for foundations of power-law distributions, the MaxEnt prescription should no doubt play a key rôle. As mentioned, MaxEnt represents a sort of "maximum honesty principle" according to which the sought distribution is the only one satisfying the observational constrains and in all other aspects being maximally "noncomitant". Otherwise, one would risk introducing false information into system's description. On the other hand, the previous discussion suggests that a statistical inference based on MaxEnt should go *beyond* Shannon's information theory in order to successfully address the ubiquitous appearance of heavy tailed distributions. This begets a very natural question; How this can be done? What element of the classical information theory should be modified so that one could still retain the desirable principle of maximal ignorance? A typical solution to this question is in a sense pedestrian; one introduces a number of *generalized* entropies that have various degrees of justification. The heavy tails may then appear (and they often do) when the MaxEnt principle is applied to such entropies. In particular, post-Shannonian developments of information and estimation theory offer various generalized measures of information to deal with such situations [153, 154]. Measures of Havrda–Charvát [122], Tsallis [311], Sharma–Mittal [289], Rényi [263, 264], Kapur [164], Landsberg–

Vedral [193, 192], Kaniadakis [162, 163] and Naudts [234] can serve as examples. In the following I will briefly discuss two most prominent examples that have represented a recurrent theme in my research, namely Tsallis–Havrda–Charvát entropy and Rényi's entropy.

1.2.7 *Tsallis–Havrda–Charvát entropy*

After Havrda and Charvát published their 1967 seminal paper in which they introduced axiomatically a new non-additive information measure called *structural α-entropy* [122], it has become obvious that other generalizations of Shannon's information measure (which differ from the venerable Rényi's generalization) are possible. However, because of the lack of an operational meaning and inherent non-additivity or non-extensivity[13] the entropy of Havrda and Charvát had laid dormant for nearly 20 years.

Inspired by multifractals, Tsallis [311] rediscovered in 1988 Havrda–Charvát's entropy. He turned its non-extensive nature into a virtue by observing that many complex systems exhibit, indeed, behavior that can be modeled by Havrda–Charvát non-extensive entropies.

$$\mathcal{S}_q(\mathcal{P}) \;=\; \frac{1}{q-1}\left(1-\sum_n^W p_n^q\right), \qquad q>0\,. \tag{1.54}$$

Entropy (1.54) is presently known as Tsallis–Havrda–Charvát (THC) entropy. In a statistical-physics community, expression (1.54) is also known under the name Tsallis' entropy.

By applying l'Hôpital's rule one can easily show that for $q=1$ the THC entropy reduces to Shannon's entropy. In addition, the THC entropy is not-additive in the sense that is satisfies the chain rule

$$\mathcal{S}_q(\mathcal{P}_1\cap\mathcal{P}_2) \;=\; \mathcal{S}_q(\mathcal{P}_1)+\mathcal{S}_q(\mathcal{P}_2|\mathcal{P}_1)+(1-q)\mathcal{S}_q(\mathcal{P}_1)\mathcal{S}_q(\mathcal{P}_2|\mathcal{P}_1)\,. \tag{1.55}$$

This is a very peculiar property — at least in the framework of classical information theory since it implies, among other things, that entropy of two independent statistical events cannot be described as a sum of two

[13] A physical observable based on a joint probability distribution is said to be *additive* if for a factorizable or separable distribution it can be written as the sum of the one-particle (or marginal) contributions. A physical observable based on the joint state of n sub-states is said to be *extensive* if it scales proportionally with n or, at least, asymptotically proportionally with n when this number becomes large. Though in most cases encountered in physics, additivity does imply extensivity (and vice versa) in general this is not the case [309]. In the following I will loosen a bit rigor and consider non-extensivity and non-additivity to be synonymous — a practice that is not uncommon in the Tsallis thermostatistics.

autonomous entropies. This makes the THC entropy difficult to fit within some operational schema. Despite the great deal of work dedicated to the formal study of the THC expression (1.55) (see, e.g., Ref. [314]), far more than the entropy $\mathcal{S}_q$ itself are in practice used ensuing MaxEnt probability distributions. In fact, a typical practical utility of Tsallis statistical paradigm is to find various fits to experimental data and limits thus values of q. On a theoretical side one is trying to use MaxEnt principle in order to impose limits on the amount by which q is likely to differ from the Shannon value of $q = 1$. For instance, when one controls in an experiment the averaged value of energy then the MaxEnt implies the entropy maximizer — "q-canonical distribution" in the form (cf Chapter 1.2.8)

$$p_n \;=\; \frac{1}{Z_q}\left[1+(q-1)\beta E_n\right]^{1/(1-q)}\,. \tag{1.56}$$

It is clear that for large spectral values of energy the tails approach a power law behavior $p_n \propto E_n^{1/(1-q)}$. This type of the THC maximizer is successfully applied in many complex systems (see, e.g., references in [315]), thought the rôle of Tsallis' inverse temperature β is as yet not fully understood. Some cases where β can be given an operational meaning are discussed in Chapter 3 in connection with superstatistics.

It needs to be highlighted that Tsallis's 1988 paper gave rise to an entirely new field known as *non-extensive* (or *Tsallis'*) *thermostatistics*. The latter has become a fertile ground for study of numerous complex dynamical systems (cf Ref. [315]). It should be also mentioned that the structure of non-extensive thermostatistics is interesting from a mathematical point of view. This is because Tsallis' thermostatics can be viewed as a deformed Shannon's information theory. Indeed, by using the basic relations from the the so called q-calculus (also known as quantum calculus) of Jackson [136], Kac [161], Cones [61], Majid [211] and others, one can introduce two basic functions: q-exponent

$$\exp_q(x) \;\equiv\; \left[1+(1-q)x\right]_+^{1/(1-q)}\,, \tag{1.57}$$

and the q-logarithm

$$\log_q(x) \;\equiv\; \frac{x^{1-q}-1}{1-q}, \qquad x>0\,. \tag{1.58}$$

The symbol $[z]_+$ is defined so that

$$[z]_+ \;=\; \begin{cases} z, \text{ if } z \geq 0\,; \\ 0, \text{ if } z < 0\,. \end{cases} \tag{1.59}$$

The previous q-functions fulfill the following important relations:

$$\begin{aligned} \log_q(\exp_q(x)) &= \exp_q(\log_q(x)) = x\,, \\ \exp_q(x)\exp_q(y) &= \exp_q(x \oplus_q y)\,, \\ \log_q(xy) &= \log_q x \oplus_q \log_q y\,. \end{aligned} \tag{1.60}$$

Here $x \oplus_q y \equiv x + y + (1-q)xy$ is the q-deformed (or Jackson's) sum. First and third equation in (1.60) allows to define the q-product and q-division as the following binary relations:

$$\begin{aligned} x \otimes_q y &= [x^{1-q} + y^{1-q} - 1]_+^{1/(1-q)}\,, \\ x \oslash_q y &= [x^{1-q} - y^{1-q} + 1]_+^{1/(1-q)}\,. \end{aligned} \tag{1.61}$$

These fulfil the consistency identities:

$$\begin{aligned} \exp_q(x+y) &= \exp_q(x) \otimes_q \exp_q(y)\,, \\ \exp_q(x-y) &= \exp_q(x) \oslash_q \exp_q(y)\,, \\ \log_q(x \otimes_q y) &= \log_q(x) + \log_q(y)\,, \\ \log_q(x \oslash_q y) &= \log_q(x) - \log_q(y)\,. \end{aligned} \tag{1.62}$$

Note that all the above q-functions satisfy the desired limiting conditions

$$\log_q(x) \overset{q\to 1}{\to} \log(x) \quad \text{and} \quad \exp_q(x) \overset{q\to 1}{\to} \exp(x)\,. \tag{1.63}$$

Also the involved q-deformed operations correctly approach the standard behavior in the $q \to 1$ limit.

The q-functions (1.60) allow to view the THC entropy as the q-deformed version of Shannon's (or Boltzmann–Gibbs's) entropy, namely

$$\begin{aligned} \mathcal{S}_q(\mathcal{P}) &= \sum_n^W p_n \log_q(1/p_n) = -\sum_n^W p_n \log_{2-q} p_n \\ &= -\sum_n^W p_n^q \log_q p_n\,. \end{aligned} \tag{1.64}$$

The analogy between the Shannon and the THC entropy can be seen from yet another point of view. Let me define the "multiplicative" Jackson q-derivative as [161]

$$D_q(f(x)) \equiv \frac{f(qx) - f(x)}{qx - x}, \quad \text{so that} \quad \lim_{q\to 1} D_q(f(x)) = \frac{df(x)}{dx}\,. \tag{1.65}$$

With this the THC entropy can be written as

$$\mathcal{S}_q(\mathcal{P}) = -\left[D_q\left(\sum_n^W p_n^x\right)\right]\Bigg|_{x=1}\,. \tag{1.66}$$

This should be compared with an analogous relation for the Shannon entropy which reads

$$H(\mathcal{P}) \;=\; -\left[\frac{d}{dx}\left(\sum_n^W p_n^x\right)\right]\Bigg|_{x=1} . \tag{1.67}$$

Since the the Jackson q-derivative is often used as a derivative on the lattice [55] one can consider the THC entropy as a natural candidate for the entropy on a discrete configuration space — lattice entropy.

It should be stressed that if only the first and second moment of p_n are experimentally (or theoretically) controlled then the THC entropy maximizers are the so-called q-Gaussian distributions [320]. Note that the q-Gaussian distribution is nothing but a (normalized) q-exponent with a quadratic argument — as could be expected by analogy with Shannon's entropy. Because of its intimate connection with the q-calculus, the nonextensive statistics appears to be strikingly mathematically coherent. This, in part, explains its popularity in a large segment of a statistical physics community.

The heavy-tailed nature of the distribution (1.56) suggests that the nonextensive thermostatistics can be conveniently used for certain strongly correlated random variables, whose correlations do not rapidly decrease with increasing distance and/or time. Since the CLT does not hold if correlations between far-ranging random variables are not negligible, one can expect that the THC entropy might serve as a tool for derivation of the generalized CLT as is the case of ordinary CLT and SE. Such a q-generalized version of the central limit theorem (q-CLT) was formulated by various authors, see, e.g., Refs. [127, 319, 320]. The essence of the proofs is, in one way or another, based a q-generalized version of Barron's information-theoretic proof [18] known from Shannon's information theory (cf also Appendix B.3). Since the definitions and lemmas entering proofs of the q-CLT make the prerequisite mathematics too lengthy, I will, for definiteness, extract only the essential points from Umarov *et al.* version of the q-CLT [319].

Let me first define the notion of q-moments.

Definition 1.1. *Let X be a continuous random variable and $f(x)$ its PDF. Let, in addition,*

$$\nu_q(f) \;\equiv\; \int_{-\infty}^{\infty} dx\, [f(x)]^q \;<\; \infty\,,$$

for certain values of $q > 0$. I define the escort (or zooming) PDF as

$$f_q(x) \;=\; \frac{[f(x)]^q}{\nu_q(f)}\,.$$

The a-th q-moment is then defined as

$$\langle X^a \rangle_q \;=\; \int_{-\infty}^{\infty} dx\, x^a f_q(x)\,,$$

in particular the q-mean is defined as

$$\mu_q \;=\; \int_{-\infty}^{\infty} dx\, x f_q(x)\,,$$

and the q-variance is defined as

$$\sigma_q^2 \;=\; \int_{-\infty}^{\infty} dx\, (x-\mu_q)^2\, f_q(x)\,.$$

With these definitions the Umarov *et al.* states the following q-CLT; Let $\{X_k\}_{k\leq N}$ be a sequence of equally distributed random variables with a finite q-mean μ_q and finite $(2q-1)$-variance σ^2_{2q-1} which are correlated according to a certain condition $\mathcal{C}_N(q)$. Then the partial sum

$$S_N \;=\; \frac{1}{D_N(q)}(X_1 + \ldots + X_N - N\mu_q)\,,$$

with $D_N(q)$ being an appropriate scaling tends in the limit $N \to \infty$ to a q-Gaussian distribution, namely

$$G_q(x) \;\propto\; \exp_q(-x^2)\,,$$

The theorem is restricted to $1 < q < 2$. The conditions $\mathcal{C}_N(q)$ are often referred to as q-independence as for $q \to 1$ it reduces to the usual condition for independent random variables.

Regardless of their potential applicability in correlated systems, the q-CLT's are notoriously difficult to operationalize. First, the q-independencies $\mathcal{C}_N(q)$ are difficult to treat analytically. In addition, so far there is not know any system (not even a model system) that would obey $\mathcal{C}_N(q)$. In the absence of examples it is difficult to see why nature would produce exactly this type of correlations among its variables. Second, the mathematics involved in the proofs, such as the q-Fourier transform or the q-independence has very peculiar features (e.g., non-uniqueness of the inverse q-Fourier transform and non-uniqueness of the q-independence) that are conceptually not very desirable. As yet, the q-CLT's represent rather interesting curiosity than a full-fledged tool.

Non-extensive thermostatistics of Tsallis has grown into a vast research discipline in a very short time. There are many interesting features that can be calculated in detail. However, enthusiasm sometimes carries its supporters too far. Promises are made concerning the ultimate generalization

beyond Boltzmann and Gibbs statistical physics. It is hard to believe that this could be the full story. In any case, the non-extensive thermostatistics provides for a new framework allowing one to find number hitherto unsuspected structures and conceptual connections in complex dynamical systems.

1.2.8 *Rényi entropy*

Out of the many possible generalizations of the concept of entropy in information theory one of the most important is Rényi's entropy (RE). RE was introduced by Hungarian mathematician and information theorist Alfred Rényi [263, 264]. Applications of RE in information theory, namely its generalization to coding theorems, were carried over by Campbel [50], Csiszár [64], Aczél [2] and others. In a physical setting was RE popularized by Kadanoff *et al* [116] and Mandelbrot [216] in connection with multifractals, and by Horodecky and Bennett in connection with multipartite entanglement. Rényi's entropy corresponds to a one-parametric class of information measures defined as [144, 263, 264]

$$\mathcal{I}_q(\mathcal{P}) \;=\; \frac{1}{1-q}\log_2\left(\sum_x p^q(x)\right), \quad q \;>\; 0\,. \tag{1.68}$$

It is also remarked that if $q \to 1$, then Rényi's entropy tends to Shannon's entropy analogously as the THC entropy. Thus the parameter $q-1$ characterizes the departure from the usual Boltzmann–Gibbs statistics or from Shannonian information theory. Shannon's entropy can be therefore considered as a measure of information of order 1. In contrast to the THC entropy the RE is additive, in fact, it satisfies the chain rule

$$\mathcal{I}_q(\mathcal{P}_1,\mathcal{P}_2) \;=\; \mathcal{I}_q(\mathcal{P}_1) \;+\; \mathcal{I}_q(\mathcal{P}_2|\mathcal{P}_1)\,. \tag{1.69}$$

In the spirit of the information theory one can interpret the previous as implying that the uncertainty (entropy) of the joint event $\mathcal{P}_1 \otimes \mathcal{P}_2$ is the uncertainty of $\mathcal{P}_1$ plus uncertainty of $\mathcal{P}_2$ when $\mathcal{P}_1$ is known. In addition, by looking at the conditional extremum of $\mathcal{I}_q(\mathcal{P})$ under the constrain $\sum_i p_i = 1$ yields that $\max_{\mathcal{P}}\mathcal{I}_q(\mathcal{P})$ is attained only for the uniform distribution $\mathcal{P} = \{1/n,\ldots,1/n\}$. The latter corresponds to a maximal ignorance about the system as any additional constraint (reflecting extra knowledge) would provide lower conditional extremum. The RE also fulfills the analogue of the second law of "thermodynamics", known from Shannon's case, i.e.

$$\mathcal{I}_q(\mathcal{B}|\mathcal{A}) \;\leq\; \mathcal{I}_q(\mathcal{B})\,. \tag{1.70}$$

Relation (1.70) says, among others, that in the course of an evolution a system looses memory of its boundary conditions — e.g., chaotic behavior, with equality if and only of $\mathcal{B}$ and $\mathcal{A}$ are independent events (i.e., knowing outcome $\mathcal{A}$ does not have any effect on the distribution of outcome $\mathcal{B}$).

For the sake of simplicity I shall consider here only the analog of canonical ensembles, where the prior information is characterized by a fixed energy expectation value. The corresponding MaxEnt distributions for $\mathcal{S}_q^{(R)} \equiv \mathcal{I}_q$ and $\mathcal{S}_q^{(THC)} \equiv \mathcal{S}_q$ can be obtained by extremizing the associated Lagrangians

$$L_q^{(R;THC)}(\mathcal{P}) \;=\; \mathcal{S}_q^{(R;THC)} - \alpha \sum_i p_i - \beta \langle H \rangle_r \,, \tag{1.71}$$

where α and β are Lagrange multipliers, the latter being the analog of the inverse temperature in natural units. The subscript r on the energy expectation value $\langle H \rangle$ distinguishes two conceptually different approaches. In information theory one typically uses the linear mean, i.e.,

$$\langle H \rangle_1 \;\equiv\; \langle H \rangle_{r=1} \;=\; \sum_i p_i E_i \,, \tag{1.72}$$

while in non-extensive thermostatistics it is customary to utilize a non-linear mean

$$\langle H \rangle_q \;\equiv\; \langle H \rangle_{r=q} \;=\; \sum_i P_i(q) E_i \,; \;\; P_i(q) \;\equiv\; \frac{p_i^q}{\sum_i p_i^q} \,. \tag{1.73}$$

The distribution $P_i(q)$ is called *escort* or *zooming* distribution and it has its origin in chaotic dynamics [26], albeit it was Rényi who first introduced it in Ref. [264]. Simple analysis reveals [19] that

$$\frac{\delta L_q^{(R)}(\mathcal{P})}{\delta p_i} \;=\; 0$$

$$\Rightarrow \; \begin{cases} p_i^{(1)} = Z_R^{-1} \left[1 - \tilde{\beta}(q-1)\Delta E_i \right]^{1/(q-1)} , & \text{for } \langle H \rangle_{r=1} \,, \\ p_i^{(2)} = Z_R^{-1} \left[1 - \beta(1-q)\Delta E_i \right]^{1/(1-q)} , & \text{for } \langle H \rangle_{r=q} \,. \end{cases} \tag{1.74}$$

Here $\tilde{\beta} = \beta/q$ and $\Delta E_i = E_i - \langle H \rangle_r$. By the same token one obtains for the THC case [19]

$$\frac{\delta L_q^{(THC)}(\mathcal{P})}{\delta p_i} \;=\; 0$$

$$\Rightarrow \; \begin{cases} p_i^{(1)} = Z_{THC}^{-1} \left[1 - \tilde{\beta}^*(q-1)\Delta E_i \right]^{1/(q-1)} , & \text{for } \langle H \rangle_{r=1} \,, \\ p_i^{(2)} = Z_{THC}^{-1} \left[1 - \beta^*(1-q)\Delta E_i \right]^{1/(1-q)} , & \text{for } \langle H \rangle_{r=q} \,, \end{cases} \tag{1.75}$$

with $\beta^* = \beta/\sum_i p_i^q$ and $\tilde{\beta}^* = \tilde{\beta}/\sum_i p_i^q$. So in contrast to (1.74), the THC MaxEnt distributions are self-referential. Generalized distributions of the form (1.74) and (1.75) are known as Tsallis distributions (or Tsallis thermostatistics distributions) and they indeed appear in one form or another in numerous statistical systems [315]. For historical reasons is $\mathcal{P}^{(1)} = \{p_i^{(1)}\}$ in (1.75) also known as the Bashkirov's 1-st version of thermostatistics, while $\mathcal{P}^{(2)} = \{p_i^{(2)}\}$ in (1.75) is called the Tsallis' 3-rd version of thermostatistics. Important feature of Tsallis distributions is that they are invariant under uniform shift of the energy spectrum since ΔE_i does not change under such a shift.

In passing, observe that the RE is related to the THC entropy through simple relations:

$$\begin{aligned} \mathcal{S}_q &= \frac{1}{q-1}\left[1 - 2^{(q-1)\mathcal{I}_q}\right], \\ \mathcal{I}_q &= \frac{1}{1-q}\log_2\left[(1-q)\mathcal{S}_q + 1\right]. \end{aligned} \tag{1.76}$$

The above monotonic relation between $\mathcal{S}_q$ and $\mathcal{I}_q$ is the reason why they have the same global (i.e., unconditional) maximizer, thought the conditional maximizer is generally different (cf Eq. (1.74) and Eq. (1.75)).

Despite the many conceptual parallels between the RE and both the SE and the THC there is as yet no generalized CLT based on the RE. It might be, however, hoped that a suitable generalization of Barron's information-theoretic proof [18] coupled with Campbell's coding theorem [50] will facilitate this task.

Let me finally stress that there exist various generalizations of Rényi entropies to the quantum mechanical setting. Most prominent among theses are Petz's quasi-entropies [254] and Renner's conditional min-, max-, and collision entropies [232, 262]. Nevertheless, the situation in the quantum setting is much less satisfactory in that these generalizations do not have any operational underpinning and, in addition, they are incompatible with each other in number of ways. For instance, whereas the classical conditional min-entropy can be naturally derived from the Rényi divergence, this does not hold for their quantum counterparts.

Appendix A

Reprinted papers on "Going Beyond B–G Statistics"

[152] P. Jizba and J. Korbel, *On the Uniqueness Theorem for Pseudo-Additive Entropies*, Entropy **19** (2017) 605.

[156] P. Jizba, J. Korbel and V. Zatloukal *Tsallis thermostatics as a statistical physics of random chains*, Phys. Rev. E **95** (2017) 022103.

[153] P. Jizba and J. Korbel, *Maximum Entropy Principle in Statistical Inference: Case for Non-Shannonian Entropies*, Phys. Rev. Lett. **122** (2019) 120601.

[154] P. Jizba and J. Korbel, *When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics*, Phys. Rev. E **101** (2020) 042126.

Appendix B

Additional notes on the central limit theorem

B.1 Proof of the CLT and the Lindeberg version of the CLT

The Central Limit Theorem was first proved by A. de Moivre for Bernoulli trials and appeared in his book, *The Doctrine of Chances*, first published in 1718. In its various reincarnations (and with varying degrees of rigor) was the CLT proved by P.-S. Laplace, A. Lyapunov and more recently by J.W. Lindeberg and P. Lévy. The CLT basically states that the normalized sum of a large number of mutually independent random variables with finite variances tends to the Normal cumulative distribution function. Here I first furnish the statement and proof of the classical univariate CLT for independent identically distributed (IID) random variables. Generalization to the multivariate CLT and for generic independent random variables with finite variances will be presented in sub-Section B.2.

Theorem B.6 (Central Limit Theorem). *Consider a sequence of IID random variables* $\{X_k\}$ *with* $k = 1, 2, \ldots, N$. *Let the common mean* $\langle X_k \rangle = \mu$ *and common variance* $\langle (X_k - \mu)^2 \rangle = \sigma^2$. *The distribution for the random variable* $\hat{S}_N = [\sum_{i=1}^{N}(X_i - \mu)]/\sqrt{N}$ *approaches in the large-N limit the Normal (or Gaussian) variable with the mean* 0 *and variance* σ^2.

Proof: Because X_k's are IID, the joint distribution is $p(\mathbf{x}) = p_N(x_1, x_2, \ldots, x_N) = p(x_1)p(x_2)\cdots p(x_N)$. The distribution of $\hat{S}_N$ is then given by integrating $p(\mathbf{x})$ over all possible values of the X_i subject to the constraint $\hat{S}_N = [\sum_{i=1}^{N}(X_i - \mu)]/\sqrt{N}$, i.e.,

$$f_N(x) \;=\; \int_{\mathbb{R}^N} d\mathbf{x}\, \delta(x - \hat{S}_N) p(\mathbf{x})\,. \tag{B.1}$$

To proceed, it is convenient to compute the characteristic function of $f_N(x)$. This can be done as follows

$$\begin{aligned}\tilde{f}_N(k) &= \int_{-\infty}^{\infty} dx\, e^{ikx} f_N(x) = \int_{\mathbb{R}^N} d\mathbf{x}\, p(\mathbf{x}) \int_{-\infty}^{\infty} dx\, e^{ikx}\, \delta(x - \hat{S}_N) \\ &= \int_{\mathbb{R}^N} d\mathbf{x}\, p(\mathbf{x})\, e^{ik\hat{S}_N} = \left[\int_{-\infty}^{\infty} dx_1 p(x_1) e^{ik(x_1-\mu)/\sqrt{N}}\right]^N \\ &= e^{-ik\mu\sqrt{N}} \left[\sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \frac{\langle x_1^n\rangle}{\sqrt{N^n}}\right]^N . \end{aligned} \tag{B.2}$$

The sum over averages can be exponentiated with the help of cumulant expansion, namely

$$\begin{aligned}\tilde{f}_N(k) &= e^{-ik\mu\sqrt{N}} \exp\left[N \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \frac{\langle x_1^n\rangle_c}{\sqrt{N^n}}\right] \\ &= \exp\left[N \sum_{n=2}^{\infty} \frac{(ik)^n}{n!} \frac{\langle x_1^n\rangle_c}{\sqrt{N^n}}\right]. \end{aligned} \tag{B.3}$$

The first few cumulants can be easily deduced by comparing the expansions in (B.2) and (B.3), namely

$$\begin{aligned}\langle x\rangle_c &= \langle x\rangle = \mu\,, \\ \langle x^2\rangle_c &= \langle (x - \langle x\rangle)^2\rangle = \langle x^2\rangle - \mu^2 \equiv \sigma^2\,, \\ \langle x^3\rangle_c &= \langle (x - \langle x\rangle)^3\rangle\,, \\ \langle x^4\rangle_c &= \langle (x - \langle x\rangle)^4\rangle - 3\langle (x - \langle x\rangle)^2\rangle^2\,. \end{aligned} \tag{B.4}$$

So we may write (B.3) as

$$\tilde{f}_N(k) = \exp\left[-\frac{k^2\sigma^2}{2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)\right]. \tag{B.5}$$

Inverse Fourier transform then yields

$$f_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right] + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{B.6}$$

Since all cumulants of X_1 beyond the second are suppressed by powers of $1/\sqrt{N}$, they vanish for $N \to \infty$. The distribution of $\hat{S}_N$ must in this limit be a Gaussian with the mean 0 and variance σ^2. □

Notes to proof: The CLT can be established under less restrictive conditions on $p(x)$ than those employed above. First, note that the above

derivation assumes that all moments exists. This assumption is, in fact, not needed. One can skip the expansion on the last-line in (B.2) and write instead $e^{-ik\mu\sqrt{N}}\left[\tilde{f}\left(\frac{k}{\sqrt{N}}\right)\right]^N$. Here, $\tilde{f}(\ldots)$ is the characteristic function related to a single random variable, say x_1. Since the characteristic function of a distribution always exists (integral converges absolutely, and hence uniformly), even when some of the moments do not. The fact that the existence of a moment of a certain order implies the existence of the corresponding derivative of the characteristic function allows to expand (B.2) to a second order in k, namely

$$\begin{aligned} & e^{-ik\mu\sqrt{N}}\left[\tilde{f}\left(\frac{k}{\sqrt{N}}\right)\right]^N \\ &= \left[\left(1+\tilde{f}'(0)\frac{k}{\sqrt{N}}+\tilde{f}''(0)\frac{k^2}{2N}+\ldots\right)\left(1-\frac{ik\mu}{\sqrt{N}}-\frac{k^2\mu^2}{2N}+\ldots\right)\right]^N \\ &= \left[1-\frac{\sigma^2k^2}{2N}+O\left(\frac{1}{N^{3/2}}\right)\right]^N \overset{N\to\infty}{\to} e^{-\sigma^2k^2/2}\,. \end{aligned} \tag{B.7}$$

Let us note that on the level of random variables $\hat{S}_N$ the CLT states nothing but that $\hat{S}_N$ *converge in distribution* (or *probability*) to a normally distributed random variable and one may write $\hat{S}_N \overset{d}{\to} Z \sim N(0,\sigma^2)$.

Furthermore, one does not need insist on identically distributed X_k's. In fact, this condition can be traded for a less restrictive condition — the so-called Lindeberg's condition. Ensuing Lindeberg's CLT is formulated as follows [201]:

Theorem B.7 (Lindeberg's Central Limit Theorem). *Consider a sequence of independent random variable* $\{X_k\}$ *with* $k=1,2,\ldots,N$. *Let the expected values are* $\langle X_k\rangle=\mu_k$ *and variance* $\langle(X_k-\mu_k)^2\rangle=\sigma_k^2$. *Let also* $\Sigma_N^2=\sum_{i=k}^N\sigma_k^2$. *If this sequence of independent random variables satisfies Lindeberg's condition:*

$$\lim_{N\to\infty}\frac{1}{\Sigma_N^2}\sum_{n=1}^{N}\left\langle(X_k-\mu_k)^2\,\chi_{\{|X_k-\mu_k|>\varepsilon\Sigma_N\}}\right\rangle \;=\; 0\,, \tag{B.8}$$

for all $\varepsilon>0$. *Here,* $\chi_{\{\ldots\}}$ *is the indicator function. Then the distribution for the random variable* $\hat{S}_N=[\sum_{i=1}^N(X_i-\mu_i)]/\Sigma_N$ *(i.e., the standardized sum) approaches in the large-*N *limit the Normal (or Gaussian) variable with the mean* 0 *and variance* 1.

NOTE: The Lindeberg condition (B.8) guarantees that the individual variances σ_k^2 are small compared to their sum Σ_N^2 in the sense that for given

$\varepsilon > 0$ for all sufficiently large N, $\sigma_k/\Sigma_N < \varepsilon$ for $k = 1, \ldots, N$. Roughly speaking, the Lindeberg condition requires that no variance in the sequence should substantially dominate, in particular, it should not be comparable with Σ_N^2. It should be stressed that the Lindeberg condition is only sufficient condition for the CLT. It can be shown that the Lindeberg condition is also necessary provided that $\sigma_N/\Sigma_N \to 0$ as $\Sigma_N \to \infty$, see, e.g. [83, 84].

B.2 Multivariate CLT and CTL for correlated variables

We have seen that the CLT is concerned with the necessary and sufficient conditions that a sum $\hat{S}_N$ of independent random variables converges to a Gaussian random variable as $N \to \infty$. The distribution function of each of the X_j's should have a finite second moment. So far I have presented only versions for independent univariate random variables. In fact, it is not difficult to generalize the CLT to multivariate random variables. Let me state as simple generalization of the Theorem B.6:

Theorem B.8 (CLT for multivariate random variables). *Consider a sequence of IID random multivariate variables* $\{\mathbf{X}_k\}$ *with* $k = 1, 2, \ldots, N$. *Let the common mean* $\langle \mathbf{X}_k \rangle = \mathbf{m}$ *and common covariance matrix* $\langle (\mathbf{X}_k - \mathbf{m})(\mathbf{X}_k - \mathbf{m})^T \rangle = \mathbf{\Sigma}$. *The distribution for the random variable* $\hat{\mathbf{S}}_N = [\sum_{i=1}^N (\mathbf{X}_i - \mathbf{m})]/\sqrt{N}$ *approaches in the large-*N *limit the Normal (or Gaussian) distribution with the mean* $\mathbf{0}$ *and the covariance matrix* $\mathbf{\Sigma}$.

NOTE: The proof is just a vector form of the proof of the CLT from the Theorem A.6. □

An adapted version of the Central Limit Theorem remains true also for sufficiently weakly correlated variables $\{X_k\}$. To this end I define the concept of *mixing sequences*.

Definition B.2 (Mixing sequence). *A sequence of random variables* $\{X_m\}$ *is said to be mixing (or strongly mixing) if and only if*

$$\lim_{k\to\infty} [\langle f(X_{n+1}, \ldots, X_{n+q}) g(X_{n+k+1}, \ldots, X_{n+k+q}) \rangle - \langle f(X_{n+1}, \ldots, X_{n+q}) \rangle \langle g(X_{n+k+1}, \ldots, X_{n+k+q}) \rangle] = 0,$$

for any two functions f *and* g *and for any* n *and* q.

In other words, a sequence is strongly mixing if and only if the two random vectors $[X_{n+1} \dots X_{n+q}]$ and $[X_{n+k+1} \dots X_{n+k+q}]$ tend to become more and more independent by increasing k (for any n and q).

Now I can state without proof a formulation of the CLT that takes into account correlations among random variables $\{X_j\}$.

Theorem B.9 (CLT for correlated variables). *Consider a stationary and mixing sequence of random variables $\{X_k\}$ with $k = 1, 2, \dots, N$, satisfying for all k $\langle X_k \rangle = \mu$, $Var[X_k] = \sigma^2$ and*

$$\lim_{N\to\infty} Var[\hat{S}_N] \;=\; \sigma^2 + 2\sum_{i=2}^{\infty} Cov[X_1, X_i] \;=\; V\,,$$

where $\hat{S}_N = [\sum_{i=1}^{N}(X_i - \mu)]/\sqrt{N}$. Then, $\hat{S}_N$ converges in distribution to the Normal random variable with the mean 0 and the variance V.

In the literature one can find a number of further generalizations of Theorem A.9, e.g, for non-Markovian but block (or "clumpy") Markovian processes [130] or for random walks on Lie groups [297]. In all the aforementioned cases one can, roughly speaking, say that the universality of the CLT holds for all systems where the random variables $\{X_k\}$ (e.g., local events) have a "sufficiently" short-range memory and/or correlations.

B.3 Information theory and the CLT

There exists also striking connection of the CLT with Shannon's information theory. This connection was contemplated by many authors, here I present the version which is due to A.R. Barron [18]:

Theorem B.10. *Let ϕ be the normal distribution $N(0, \sigma^2)$ and let $\{X_k\}$ be a sequence of mutually independent random variables with a common distribution. Suppose that X_k (for all k) have the finite expected value μ and finite variance σ^2. Denote by g_N the probability density function of the random variable $\hat{S}_N = [\sum_{i=1}^{N}(X_i - \mu)]/\sqrt{N}$. The relative Shannon's entropy converges to zero:*

$$\lim_{N\to\infty} D(g_N||\phi) \;=\; 0\,,$$

if and only if $D(g_N||\phi)$ is finite for some N.

NOTE: The relative Shannon's entropy $D(f||g)$ is also known as the Kullback–Leibler distance (or divergence) between two probability distributions f and g. In statistics, the entropy $D(f||g)$ is the measure of inefficiency

of assuming that a systems is described by the distribution g when the true distribution is f, see, e.g. [62]. The key property of the relative entropy is that it is always *non-negative* and is *zero* if and only if $f(x) = g(x)$ for all x. This is a consequence of Jensen's inequality and it is known as Gibbs inequality [141] (see also Appendix D.2). If one knows, for instance, the true distribution g_N, then one could construct a code with average description length given by Shannon's entropy $H(g_N)$. If, instead, one used the code based on the Gaussian distribution ϕ, one would need $H(g_N) + D(g_N||\phi)$ bits on the average to describe the random variable. Barron's version of the CLT ensures that both codes would have in the large N limit the same average length. □

We have seen that the hallmark of the CLT is the independence of the cumulative distribution on details of the single event distribution. This is akin to the universality hypothesis of critical phenomena in which short-range details of a system do not affect large-scale properties [295]. This parallel goes even further. The non-Gaussian fix points in theory of critical phenomena can be associated with another class of attractive universal distributions — Lévy stable distributions (of which the Gaussian distribution is a special case).

B.4 Asymptotic behavior of the symmetric Lévy stable distributions with $\beta = 0$

Among the most prominent class of Lévy stable distributions are symmetric (i.e., $\gamma = 0$) distributions $L_{\alpha,0}(x)$. The problem of estimating the tail index (as well as other parameters) is in these cases severely hampered by the lack of known closed-form density functions for all but a few members of the $L_{\alpha,0}(x)$ family. Apart from the Gaussian ($\alpha = 2$) and Lorentz–Cauchy ($\alpha = 1$) distributions the explicit form of $L_{\alpha,0}(x)$ cannot be phrased via simple functions. Fortunately, their *asymptotic* tail behavior (which is important for the generalized CLT) can be however extracted relatively easily. To this end I write

$$\begin{aligned} L_{\alpha,0}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi\, \exp[-ix\xi - c|\xi|^{\alpha}] \\ &= \frac{1}{\pi}\int_{0}^{\infty} d\xi\, e^{-c|\xi|^{\alpha}}\, \cos(x\xi) \;=\; L_{\alpha,0}(|x|)\,. \end{aligned} \tag{B.9}$$

Before proceeding we observe that upon setting $|x| = \varepsilon^{-1}$ and $|x|\xi = \tau$ in (B.9) we obtain

$$L_{\alpha,0}(x) \;=\; \frac{\varepsilon}{\pi}\int_0^\infty d\tau\, \exp[-c(\varepsilon\tau)^\alpha]\,\cos(\tau)\,. \tag{B.10}$$

If I expand $\exp[-c(\varepsilon\tau)^\alpha]$ and formally exchange summation with integration, $L_{\alpha,0}(x)$ becomes

$$L_{\alpha,0}(x) \;=\; \frac{1}{\pi}\sum_{k=0}^\infty \frac{(-c)^k\varepsilon^{\alpha k+1}}{k!}\int_0^\infty d\tau\,\tau^{\alpha k}\,\cos(\tau)\,. \tag{B.11}$$

Unfortunately the integral inside the sum is defined only for $\Re(\alpha k) \in (-1,0)$. Since our interest is in $\alpha \in (0,2)$ we should in some way extend the definition region of the integration. This can be done by various means. In present approach I introduce into the integral a spurious exponentially damping term, namely

$$\int_0^\infty d\tau\,\tau^{\alpha k}\,\cos(\tau) \;=\; \lim_{\epsilon\to 0_+}\int_0^\infty d\tau\, e^{-\epsilon\tau}\tau^{\alpha k}\,\cos(\tau)\,. \tag{B.12}$$

The right-hand-side is now defined for any value of $k\alpha$ provided $\Re(\alpha k) > -1$ and $\epsilon > 0$. An explicit integration and successive limit give

$$\int_0^\infty d\tau\,\tau^{\alpha k}\,\cos(\tau) \;\stackrel{\text{def.}}{=}\; -\Gamma(k\alpha+1)\sin\left(\frac{k\pi\alpha}{2}\right). \tag{B.13}$$

Prescription (B.12) basically represents an analytical extension of the integral. In fact, while the left-hand-side of (B.13) justifies the right-hand-side only for $\Re(\alpha k) \in (-1,0)$, the latter is analytic for any $\alpha k \neq -1,-2,-3,\cdots$. Since (B.10) is analytic for all α, I can directly work with an analytically extended version of the integral (B.13). With this proviso I can write

$$\begin{aligned} L_{\alpha,0}(x) \;&=\; -\frac{1}{\pi}\sum_{k=0}^\infty \frac{(-c)^k}{k!}\,\varepsilon^{\alpha k+1}\Gamma(k\alpha+1)\sin\left(\frac{k\pi\alpha}{2}\right) \\ &=\; -\frac{1}{\pi}\sum_{k=1}^\infty \frac{(-c)^k}{k!}\frac{\Gamma(\alpha k+1)}{|x|^{\alpha k+1}}\sin\left(\frac{k\pi\alpha}{2}\right). \end{aligned} \tag{B.14}$$

From this one can obtain the two-term asymptotic approximation of $L_{\alpha,0}(x)$ for large $|x|$ as a function of the parameter α,

$$L_{\alpha,0}(x) \;\simeq\; \frac{c\Gamma(\alpha+1)}{\pi|x|^{\alpha+1}}\sin\left(\frac{\pi\alpha}{2}\right) \;-\; \frac{c^2\Gamma(2\alpha+1)}{\pi|x|^{2\alpha+1}}\sin(\pi\alpha)\,. \tag{B.15}$$

So, when $\alpha \neq 2$ then $L_{\alpha,0}(x)$ has a long inverse power tail.

Note that for $\alpha > 2$ the distribution $L_{\alpha,0}(x)$ is not proper since it may, for some values of x, become negative. Result (B.14) coincides with the result of Humbert [133] who obtained the very same answer by different means.

In general, one can make a similar analysis for general Lévy distribution $L_{\alpha,\beta}$. The leading-order power law tail will stay the same as in the case of $L_{\alpha,0}$. The explicit form of the leading order reads (cf. eg. Ref. [239])

$$L_{\alpha,\beta}(x) \;\simeq\; \frac{c(1+\mathrm{sgn}(x)\beta)\Gamma(\alpha+1)}{\pi|x|^{\alpha+1}} \sin\left(\frac{\pi\alpha}{2}\right). \tag{B.16}$$

B.5 Probability space as a measure space

A triple $(X, \mathcal{F}, \mu)$ is called a *measure space*. Here, X is a non-empty set, $\mathcal{F}$ is a σ-algebra on the set X, and μ is a measure on $\mathcal{F}$. By a σ-algebra (or also σ-field) on X is meant a collection $\mathcal{F}$ of all subsets of X that includes the empty subset, is closed under complement, and is closed under countable unions and countable intersections. A simple example of measure spaces is the n-dimensional Euclidean space, with the Borel σ-algebra on $\mathbb{R}^n$ (i.e. $\mathcal{B}(\mathbb{R}^n)$) and with Lebesgue measure. This also serves as a canonical measure space in integral calculus.

An important example of a measure space is a *probability space*. A probability space is a measure space where the triple is typically denoted as $(\Omega, \mathcal{F}, P)$ and $P(\Omega) = 1$. Here, Ω is a sample space (i.e., the set of all possible outcomes of a probabilistic experiment), the σ-algebra $\mathcal{F}$ is the set of all subsets of Ω that are considered as events, and the probability measure P is a function that associates a probability to each of the events belonging to $\mathcal{F}$.

The actual probability measure is a set of functions $P(\ldots)$ that assigns to every event $E \in \mathcal{F}$ a number $0 \leq P(E) \leq 1$ called the probability of event E. This is done via *Kolmogorov axioms*:

Axiom 1 (Kolmogorov axioms).

(1) $P(E) \geq 0$, $P(E) \in \mathbb{R}$ *for all* $E \in \mathcal{F}$.
(2) $P(\Omega) = 1$.
(3) Assumption of σ*-additivity. Any countable sequence of disjoint sets (mutually exclusive events, i.e.* $E_i \cap E_j = \emptyset$*, for* $i \neq j$*)* $E_1, E_2, \ldots$ *satisfies*

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) \;=\; \sum_{i=1}^{\infty} P(E_i)\,.$$

Note: Some authors consider merely finitely additive probability spaces (i.e., index i in the 3rd axiom is bounded from above), in which case one just needs an algebra of sets, rather than a σ-algebra. On the other hand, infinite additivity enables one to deal rigorously with limits and countable unions.

Three above axioms are sufficient to establish other useful rules for calculating probabilities. For instance [85]

- $P(\emptyset) = 0$.
- $P(E \cap F^c) = P(E) - P(E \cap F)$ *where* F^c *is the complement of a set* F.
- $P(E) = 1 - P(E^c)$.
- *"Monotonicity property". If* $E \subseteq F$ *then* $P(E) \leq P(F)$. *This directly implies that* $0 \leq P(E) \leq 1$ *for all* $E \in \mathcal{F}$.

To illustrate the use of Kolmogorov axioms let me consider, for instance, the experiment consisting of throwing a dice once. The sample space Ω are possible outcomes (the number of dots appearing on the upward-facing side of the dice), i.e., $\Omega = \{1, 2, 3, 4, 5, 6\}$. The σ-algebra of events consists of

$$\sum_{n=0}^{6} \binom{6}{n} \;=\; 2^6\,, \tag{B.17}$$

events. Some typical events from $\mathcal{F}$ are: $\emptyset, \Omega, \{1\}, \{1, 2\}, \{1, 2, 3\}, \{1, 3, 6\}$ and $\{2, 3, 4, 6\}$. The measure/probability assignment is done by affiliating probabilities to the elementary events $\{i\}$ so that $P(\{i\}) = 1/6$ for $i = 1, \ldots, 6$. All remaining probabilities can be computed from the Kolmogorov axioms and the assumed probabilities of elementary events. For instance, with $A \equiv \{1\}$ and $B \equiv \{3, 5\}$, I can write $P(A \cup B) = P(A) + P(B)$ since A and B are mutually exclusive events. In addition, $P(B) = P(\{3\}) + P(\{5\}) = 2/6$. Consequently, $P(A \cup B) = P(\{1, 3, 5\}) = 1/6 + 2/6 = 1/2$.

Note: The Kolmogorovian axiomatic approach for the assignment of probabilities (which parallels with a measure theory), is of major conceptual importance, and provides the mathematical foundation for *all* standard uses of probability theory, including the whole of classical statistical physics. In the Kolmogorovian approach, the probabilities are specified as measures on sets, and if the sets are subsets of a multi-dimensional space they have the properties of volume. In fact, the Kolmogorovian probability of a subset E_i is given as a relative volume with respect to the entire set's volume Ω by $\mathrm{Vol}(E_i)/\mathrm{Vol}(\Omega)$. The probabilities thus obtained behave like relative frequencies, and indeed two disjoint (relative) volumes can be added to give the relative volume of the union.

The key feature of probabilities used in quantum theory is that they do *not* arise in this way but rather from something quite different, namely from the (n-dimensional) *Pythagorean theorem*. Indeed, the important ingredient of the Kolmogorov probability theory is the additivity property of the measure for mutually exclusive events, while in quantum theory such additivity holds only for probability amplitudes (which do not need to add up to unity).

In Fig. B.1 I denote with $\mathbf{x}$ and $\mathbf{y}$ two eigenvectors of some two-state observable X which form a 2-dimensional orthonormal basis for a 2-dimensional space. The projection of any state $\mathbf{c}$ onto $\mathbf{x}$ is a, while pro-

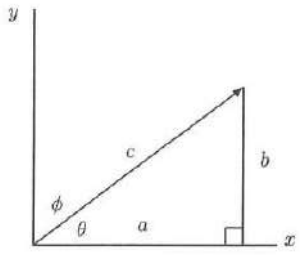


Fig. B.1: The Pythagoras theorem in 2 dimensions.

jection to $\mathbf{y}$ is b. By setting $c \equiv |\mathbf{c}|$, one has (by the Pythagoras theorem) in two dimensions $a^2 + b^2 = c^2$, which implies that $(a/c)^2 + (b/c)^2 = 1$. This is just $\cos^2\theta + \sin^2\theta = 1$ or equivalently $\cos^2\theta + \cos^2\phi = 1$ where θ and $\phi = \pi/2 - \theta$ are angles between the vector $\mathbf{c}$ and the axes x and y, respectively.

This gives a way of interpreting a probability of observing the result of an observable X with two outcomes a_1, a_2, where a_1 is the eigenvalue corresponding to the eigenvector $\mathbf{x}$, and a_2 is the eigenvalue corresponding to the eigenvector $\mathbf{y}$. The state of the system is represented by the vector $\mathbf{c}$, and if normalised its length is $c = 1$. So, one can interpret p_1 to be the probability that one gets a_1 for observable X given the state $\mathbf{c}$, where $p_1 = P(X = a_1|\mathbf{c}) = \cos^2\theta$ and $p_2 = P(X = a_2|\mathbf{c}) = \cos^2\phi$. The state $\mathbf{c}$ can vary throughout the space but $p_1 + p_2 = 1$ for any $\mathbf{c}$. In particular $p_1 = 1$ and $p_2 = 0$, if $\mathbf{c}$ lies along $\mathbf{x}$ and vice versa if $\mathbf{c}$ lies along direction $\mathbf{y}$.

This reasoning can be easily extended to higher dimensions. For instance, in 3-dimensional state space the generalized Pythagoras theorem reads: $\cos^2\theta_1 + \cos^2\theta_2 + \cos^2\theta_3 = 1$, where θ_i represent the angles made by an unite state vector $\mathbf{c}$ with three eigenvectors of some three-state observable X. This fact is known, for instance from crystallography (theory of Miller indices) [174]. Above construction constitutes the essence of the non-classical (non-Kolmogorovian) probability framework employed quantum mechanics (though using vectors whose components are complex, rather than real numbers). As noted in Ref. [134]: *... It is salutary though that the heart of the radically different natures of classical and quantum probability is just the difference between numbers obtained from ratios of volumes, and numbers that come from Pythagoras theorem!*

B.6 Infinitely divisible variables and distributions

Concept of an infinitely divisibility (ID) was introduced in 1929 by Bruno de Finetti [66] as a mean for the study of general stochastic processes with stationary independent increments. The connection of ID distributions with the generalized CLT was highlighted by Khintchine [171] and by Kolmogorov and Gnedenko [106] who proved that the limit distribution functions from the (generalized) CLT must belong to the class of infinitely divisible distributions. Opposite is, however, not true.

Roughly speaking, one can say that a distribution function is ID if it can be expressed as the probability distribution of the sum of an arbitrary number of independent and identically distributed random variables. More rigorously

Definition B.3 (Infinitely divisible variables and distributions). *A random vector X in $\mathbb{R}^d$ (or its distribution) is said to be infinitely divisi-*

ble if for every $n \geq 1$ there exist independent identically distributed random vectors $Y_{n1}, \ldots, Y_{nn}$ in $\mathbb{R}^d$ (possibly on a different probability space) such that

$$X \stackrel{d}{=} Y_{n1} + \ldots + Y_{nn} \,.$$

Equivalently, a probability density function p on $\mathbb{R}^d$ is ID if an only if for every $n \geq 1$ there exists a probability density function p_n on $\mathbb{R}^d$ such that

$$p \;=\; \underbrace{p_n * \cdots * p_n}_{n} \;=\; [p_n]^{*n} \,.$$

Consequently, the characteristic function $\varphi(p)$ associated with $p(x)$ can be written for every $n \geq 1$ as the n-th power of some characteristic function $\varphi_n(q)$. So, in particular, for any $n \geq 1$ one requires that $\varphi(p) = [\varphi_n(p)]^n$ with (i) $\varphi_n(0) = 1$ and (ii) $\varphi_n(p)$ is continuous for all p from the definition region.

Normal (i.e., *Gauss*), *Gamma* and *Poisson* distributions in $\mathbb{R}^1$ may serve as typical examples of the ID distributions. Indeed, the Gaussian distribution (of the mean μ and variance σ^2) has the characteristic function

$$\varphi(p) \;=\; \exp\left[i\mu p - \frac{\sigma^2}{2}p^2\right], \tag{B.18}$$

so that

$$\varphi_n(p) \;=\; \exp\left[i\frac{\mu p}{n} - \frac{\sigma^2}{2n}p^2\right]. \tag{B.19}$$

Analogously, the Gamma distribution (of the mean α/β and variance α/β^2) has the characteristic function

$$\varphi(p) \;=\; \left[1 - i\frac{p}{\beta}\right]^{-\alpha}, \tag{B.20}$$

so that

$$\varphi_n(p) \;=\; \left[1 - i\frac{p}{\beta}\right]^{-\alpha/n}. \tag{B.21}$$

Finally, the Poisson distribution (with mean $a + \lambda h$ and variance $h^2\lambda$) has the PDF

$$\varrho(y) \;=\; \sum_{m=1}^{\infty} \frac{\lambda^m}{m!}\, e^{-\lambda}\, \delta(y - a - mh) \,. \tag{B.22}$$

The characteristic function is

$$\varphi(p) \;=\; \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}\, e^{-\lambda}\, e^{ip(a+mh)} \;=\; \exp\left[iap + \lambda\left(e^{iph} - 1\right)\right], \qquad \text{(B.23)}$$

so that

$$\varphi_n(p) \;=\; \exp\left[ia\frac{p}{n} + \frac{\lambda}{n}\left(e^{iph} - 1\right)\right]. \qquad \text{(B.24)}$$

On the other hand, any continuous uniform distribution

$$P(x) \;=\; \begin{cases} 1/2L & \text{for } -L \;\leq\; x \;\leq\; L\,, \\ 0 & \text{for } x \;>\; L \text{ or } x \;<\; -L\,, \end{cases} \qquad \text{(B.25)}$$

has the characteristic function

$$\varphi(p) \;=\; \frac{\sin(pL)}{pL}\,. \qquad \text{(B.26)}$$

In this case the generic n-th root does not represent a genuine characteristic function. This is a direct consequence of the fact that the characteristic function $\varphi(p)$ has zero(s) [277]. Quite generally, no probability density (except Dirac δ function) with bounded support can provide ID distributions [277]. The second Lévy–Khintchine theorem or Lévy–Khintchine formula gives a structural form of cumulants of ID distributions [84, 106, 171, 196, 277].

Theorem B.11 (second Lévy–Khintchine theorem). *Let X be an ID random vector X in $\mathbb{R}^d$, let the ensuing probability density function be $L(x)$ and the characteristic function*

$$\tilde{L}(\xi) \;=\; \int_{\mathbb{R}^d} d^d x\, L(x)\, e^{i\xi\cdot x}\,, \qquad \text{(B.27)}$$

then there exits a unique triplet (b, Σ, μ) such that

$$\ln \tilde{L}(\xi) \;=\; ib\cdot\xi - \frac{1}{2}\xi\cdot(\Sigma\xi) + \int_{\mathbb{R}^d} \mu(dx)\left[e^{i\xi\cdot x} - 1 - i\xi\cdot\tau(x)\right], \qquad \text{(B.28)}$$

where $\xi, b \in \mathbb{R}^d$, Σ is a nonnegative definite $d\times d$ matrix and μ is a measure on $\mathbb{R}^d$ with $\mu(\{0\}) = 0$ and $\int_{\mathbb{R}^d}(\|x\|^2 \wedge 1)\mu(dx) < \infty$. $\tau : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a fixed bounded measurable function such that $\lim_{x\to 0}\|x\|^{-2}[\tau(x) - x] = 0$.

The matrix Σ is called a Gaussian covariance matrix, μ a Lévy measure, and τ a truncation function. Conversely, given a triple (b, Σ, μ) and τ as above, there exists an ID random vector X satisfying (B.28).

One typically writes $X \sim ID(b, \Sigma, \mu)$.

The proof of the Lévy–Khintchine characterisation of ID random variables is quite lengthy and I avoid it here. I can refer the interested reader to Refs. [204, 277] for more details. Despite its key importance, the Lévy–Khintchine canonical formula is rather cumbersome to use in practical applications and, in fact, there are many distributions where it is notoriously hard to prove ID. Examples include *Student's t-distribution*, *Pareto distribution*, *Log-normal distribution* or *Weibull distribution*. It is worth mentioning that proving or disproving infinite divisibility of certain distributions is often a very tedious task and special techniques may be needed for particular special problems.

In the class of ID distributions there is an important subclass of so-called stable processes (see Chapter 1.1.3 and Appendix B.4). Correspondingly, a special rôle belongs to the class of stable laws a stable random variables. The class of *stable* random variables and distributions, by definition, corresponds to the case when the random variables $Y_{n1}, \ldots, Y_{nn}$, in Definition A.3 can be constructed by means of a special procedure from a sequence $X_1, \ldots, X_n$ of independent and identically distributed random variables

$$Y_{nk} \;=\; a_n X_k \;+\; b_n \,, \tag{B.29}$$

where $a_n > 0$ and $b_n \in \mathbb{R}$ are suitable constants (see Chapter 1.1.3). A remarkable result of Lévy and Khintchine states that for a stable random variable the logarithm of characteristic function $\tilde{L}(\xi)$ is given by the first Lévy–Khintchine theorem (cf. Theorem 1.2). From Theorem 1.3 we have seen that $a_n = 1/n^{1/\alpha}$ for some $0 < \alpha \leq 2$ (particular α defines class of α-stable distributions). In the case of $b_n = 0$ a stable distribution is called a strictly α-stable distribution.

Let me finally mention that the collection of all ID distributions is in a one-to-one correspondence to Lévy processes which are heavily used in the financial literature [45, 222], see Chapter 3.

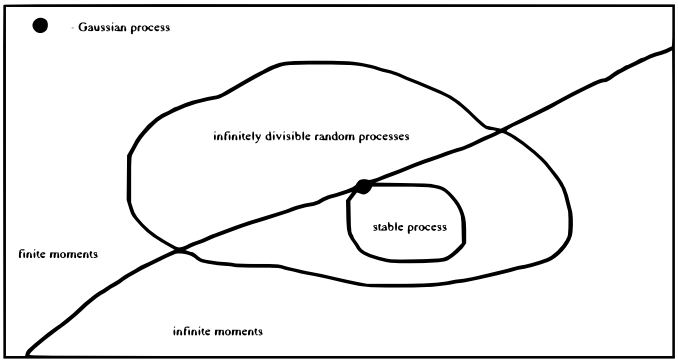


Fig. B.2: Classes of random processes associated with discussed distributions. Among distributions that have finite moments and are not infinitely divisible belong, e.g. distributions which have densities with bounded support. Among distributions that have infinite first/second moment and are not infinitely divisible belong, e.g. generic mixtures of Lévy stable distributions.

Chapter 2

Statistics of Rényi

In Section 1.2.8 I have provided some incentives for Rényi entropy. In the present chapter I discuss Rényi entropy in more detail. In particular, I show that the concept of Rényi entropy (and associated differential Rényi entropy) can be conveniently applied in theory of multifractals, in financial markets and in quantum mechanics.

2.1 Rényi enatropy as information measure

In the previous chapter I have introduced two important entropic measures with non-gaussian maximizers, namely Rényi and Tsallis entropies. A brief comparison between Chapters 1.2.7 and 1.2.8 shows that maximizers of both entropies are formally identical provided the same constraints are employed. Natural question thus arises: Why do we need RE if the associated maximizer coincides with that of THC entropy? In particular, since THC entropy is simpler to handle due to its intimate connection with the q-calculus, it would seem only natural to employ THC instead of RE. In fact, the aforementioned link with the q-calculus is one of the major reasons for a widespread use of the non-extensive statistics of Tsallis. In this chapter I will put forward various motivations in favour of RE. The selected topics are based on my research and hence inevitably reflect my personal prejudices and biases towards RE. I will briefly return to THC entropy in Section 2.4.3.6

To proceed, let me first make a brief comparison between Rényi and Tsallis entropies.

(1) $H(\mathcal{P}) \leq \mathcal{I}_q(\mathcal{P}) \leq \mathcal{S}_q(\mathcal{P})$ for $0 < q \leq 1$,

$$\mathcal{S}_q(\mathcal{P}) \;\leq\; \mathcal{I}_q(\mathcal{P}) \;\leq\; H(\mathcal{P}) \quad \text{for } q \;\geq\; 1.$$

(2) $\mathcal{I}_q$ is *additive* (or *extensive*) while $\mathcal{S}_q$ is *superadditive* for $q < 1$ and *subadditive* for $q > 1$, i.e. $\mathcal{S}_q$ exhibits a *non-extensive* behavior.

(3) $\mathcal{S}_q$ is *concave* for $q > 0$ while $\mathcal{I}_q$ is *concave* only for $0 < q \leq 1$ (irrespective of the dimensionality of the sample space). Convexity/concavity properties of $\mathcal{I}_q$ for $q > 1$ depend on dimensionality of probabilistic vectors. For instance, the binary RE is concave for $q \in [0; 2]$.

(4) Both $\mathcal{S}_q$ and $\mathcal{I}_q$ are *Schur-concave* for all $q > 0$.

(5) $\mathcal{I}_q \;=\; \log_2[(1-q)\mathcal{S}_q+1]/(1-q)$, i.e. $\mathcal{I}_q$ and $\mathcal{S}_q$ are monotonic functions of each other and hence they have *identical maximizer* under identical constraints.

(6) $\lim_{q\to 1} \mathcal{I}_q \;=\; \lim_{q\to 1} \mathcal{S}_q \;=\; H$.

(7) $\mathcal{I}_q$ has operational meaning given by *coding theorems* while for $\mathcal{S}_q$ no such operational meaning exists yet.

It is mainly the last point, which singles out RE for applications in (quantum) information theory. In fact, information-theory based entropies (i.e., information measures) are primarily important because there exist various coding theorems which endow them with an operational (that is, experimental determinable) meaning, and not because of intuitively pleasing aspects of their definitions. While coding theorems do exist both for the Shannon entropy and the Rényi entropy, there are (as yet) no such theorems for THC entropy. In order to understand the core idea behind the RE-based coding theorem let me state first the key inequality of information theory, namely the so-called Kraft–McMillan (inequality (KMI):

Theorem 2.1 (Kraft–McMillan inequality). *Let* $\{x_i\}_{i=1}^N$ *be source symbols. These can be encoded into a uniquely decodable code over the alphabet of size* D *with codeword lengths* $\{\ell_i\}_{i=1}^N$ *if and only if*

$$\sum_{i=1}^{N} D^{-\ell_i} \;\leq\; 1\,.$$

So, for instance, when I wish to encode source symbols $\{x_i\}_{i=1}^4 = \{a, b, c, d\}$ into a binary alphabet $\{0, 1\}$ I might employ the coding scheme: $a \to 0$, $b \to 10$, $c \to 110$ and $d \to 111$. Since in this case $D = 2$ and the

codeword lengths $\{\ell_i\}_{i=1}^4$ are $\{1,2,3,3\}$ I obtain that

$$\sum_{i=1}^{4} 2^{-\ell_i} \;=\; 2^{-1}+2^{-2}+2^{-3}+2^{-3} \;=\; 1\,, \tag{2.1}$$

which saturates the KMI[1]. The proof of the KMI can be found in various textbooks, cf. e.g., Ref. [8].

From Theorem 2.1 we see that the KMI specifies the fundamental constraint on the lengths of the codewords. When codeword lengths are quantified via single parameter — an expected length $\langle\ell\rangle = \sum_{i=1}^{N} p_i\ell_i$, then the KMI can be used to show that any uniquely decodable code must have $\langle\ell\rangle$ larger than or equal to Shannon's entropy. This is the essence of Shannon's coding theorem. In fact, there are various coding algorithms that saturate Shannon's bound $\langle\ell\rangle = H(\mathcal{P})$ subject to the KMI. The most familiar is the so-called Huffman coding algorithm [132].

It is, however, possible to satisfy the KMI with another types of coding schemes where different quantifiers of codewords lengths are employed. The most prominent coding theorem among these is Campbell's coding theorem (CCT) [50]:

Theorem 2.2 (Campbell's coding theorem). *Let $\{p_i\}_{i=1}^N$ be the probabilities of N input symbols $\{x_i\}_{i=1}^N$ which we wish to encode. Suppose there is an alphabet of D symbols into which the input symbols are to be encoded. Let x_k be represented by a sequence of ℓ_k characters from the alphabet satisfying the KM inequality. Then*

$$\mathcal{I}_q \;\leq\; L_\beta < \mathcal{I}_q + 1$$

where a code length of order β with $\beta \in (0,\infty)$ is

$$L_\beta \;=\; \frac{1}{\beta}\log_D\left(\sum_{i=1}^{N} p_i D^{\beta\ell_i}\right)$$

and $\beta = (1-q)/q$.

So, while in the Shannon case one can view the *cost* of a code-word as a linear function of the length, in the Rényi case the cost of a code-word is an exponential function of its length [50] with q playing the role of a cost parameter. Campbell's theorem then basically states that the optimal code for a noiseless cannel has a minimal cost out of all codes with a given cost function.

[1]This coding scheme is known as a *prefix code*. Any coding scheme where no codeword is a prefix of any other codewords is called a prefix code.

According to the CCT RE corresponds to the averaged number of bits needed to optimally encode the discrete source X with the probability $\mathcal{P}(X)$, provided that the codeword-lengths are exponentially weighted (or priced). This exponential weighting is also known as a Kolmogorov–Nagumo averaging. While the linear averaging is given by

$$\langle X \rangle \;=\; \sum_{x \in X} p(x) x \,, \tag{2.2}$$

the exponential weighting is defined as

$$\langle X \rangle_{\rm exp} \;=\; \phi^{-1} \left(\sum_{x \in X} p(x) \phi(x) \right) , \tag{2.3}$$

with $\phi(x) = D^{\beta x}$. In our case, $p(x) = p(\ell_i) \equiv p_i$ and $x = \ell_i$. The $\beta = (1-q)/q$ factor is also known as the Campbell exponent. It is easy to check that the equality in Campbell's lower-bound inequality is achieved by choosing the ℓ_i such that

$$D^{-\ell_i} \;=\; \frac{p_i^q}{\sum_{j=1}^n p_j^q} \,, \tag{2.4}$$

or equivalently

$$\ell_i \;=\; -q \log_D p_i \;+\; (1-q) \mathcal{I}_q(\mathcal{P}) \,. \tag{2.5}$$

Clearly, the individual lengths ℓ_i obtained through the saturation of Campbell's lower-bound inequality can be made smaller than the ensuing Shannons' lengths $\ell_i = -\log_D p_i$, particularly for small p_i by selecting a sufficiently small value of q. This is a simple consequence of the fact the inequality $\mathcal{I}_q(\mathcal{P}) \leq -\log_D p_i$ can be satisfied for any $p_i < 1/N$ by a sufficiently small q. So, Campbell's coding procedure effectively penalizes longer codewords and supplies a code different from Shannon's code, with generally shorter codewords associated with lower probabilities. The codewords are the shorter the lower Rényi's parameter q or, equivalently, the higher price factor β.

In this connection I should also point out that Campbell's coding theorem for RE is equivalent to Shannon's coding theorem for SE provided one uses instead of p_i the escort distribution [26, 32]:

$$\varrho(q, \ell_i) \;=\; \frac{p_i^q}{\sum_{j=1}^N p_j^q} \,. \tag{2.6}$$

Indeed, in order to see this let me define the average code-word length with the help of escort distribution, i.e.

$$\langle \ell \rangle_q \;=\; \sum_{k=1}^N \varrho(q, \ell_k) \ell_k \,. \tag{2.7}$$

p_i	$\beta = 9$ $(q = 1/10)$	$\beta = 0.5$ $(q = 2/3)$	$\beta = 0$ $(q = 1)$
0.3	01	11	0
0.38	00	10	10
0.2	111	01	111
0.1	110	001	1101
0.01	101	0001	11001
0.0	100	0000	11000

Table 2.1: Example of a lossless data compression algorithm — Huffman coder, for 6 source symbols with probabilities $\{p_i\}_{i=1}^6$ and the binary encoding alphabet $\{0, 1\}$ $(D = 2)$. Three different price factors $\beta = (1 - q)/q$ are depicted.

Shannon's coding theorem which would employ $\varrho_q = \{\varrho(q, \ell_k)\}_{k=1}^N$ instead of $\mathcal{P} = \{p_k\}_{k=1}^N$ would imply Shannon's inequality

$$H(\varrho) \;=\; \mathcal{I}_1 \;\leq\; \langle \ell \rangle_q \;\leq\; H(\varrho) \;+\; 1\,. \tag{2.8}$$

It is easy to check that the equality in Shannon's lower-bound inequality is achieved by choosing the ℓ_i such that

$$\ell_i \;=\; -\log_D \varrho(q, \ell_i) \;=\; -q \log_D p_i \;+\; (1 - q)\mathcal{I}_q(\mathcal{P})\,, \tag{2.9}$$

which is nothing but the Campbell result (2.5). So, in practical applications one can employ the Campbell coding theorem in such a way that standard Shannonian coding algorithms (such as Huffman's one) are feeded with the escort distribution ϱ_q instead of the original source distribution $\mathcal{P}$. As a simple example, I use in Tab. 2.1 a standard Huffman algorithm[2] to encode 6 source symbols with given occurrence probabilities into a binary alphabet. The corresponding codes are generated for 3 distinct price factors β corresponding to $q = 1$ (Shannon's coding), $q = 2/3$ and $q = 1/10$.

In passing one can note that the spacial case $q = 2$ was studied already prior Rényi's work by many authors, see, e.g., Refs. [16, 265] in connection with the *correlation dimension*. Among other information-theoretic applications of RE I can mention, e.g., Csiszár, Campbell and Ahlswede–Cai works on block coding [3, 51, 63] or Bennet *at al.* works on use of RE in quantum cryptography [31, 208, 224].

[2]There is a number of online Huffman coders, see, e.g.: https://planetcalc.com/2481/

2.2 World according to Rényi

Among the most prominent applications of RE is the application in fractal and multifractal systems. The purpose of this section is to stress this rôle and provide some related discussion. In my exposition I will closely follow Refs. [143–145].

2.2.1 *Fractals and Multifractals — brief introduction*

In the previous section we have seen that Rényi's q parameter quantifies the cost of a code-word in Campbell's coding theorem. Another area where the q-parameter has a well defined meaning is the theory of multifractals. In particular, there the q-parameter corresponds to the so-called *multifractal spectral dimension*. To clarify this connection, I start first with a brief introduction into theory of fractals.

Fractals typically appear in study of geometrical properties of sets and are characterized by property of self-similarity (i.e., form invariance under magnification) and by non-integer dimension D_H — Hausdorff or fractal or Mandelbrot dimension [80, 81]. Fractal patterns appear not just in the shape of coastlines (as depicted in its seminal work by Mandelbrot [214]) and in the decorative designs generated by innumerable computer programs, but they are very real. They appear in the distribution of galaxies throughout the cosmos, in biological systems but also in the price changes of securities. Some examples typical fractals with their respective fractal dimensions are depicted in Fig. 2.1:

(a) (b) (c) (d) (e)

Fig. 2.1: Some examples of real-world fractal systems: Brownian motion in $x-t$ plane ($D_H = 1.5$), coast of Norway ($D_H = 1.52$), Brownian motion in $x-y$ plane ($D_H = 2$), distribution of galaxy clusters ($D_H \sim 2$), brain ($D_H = 2.79$)

The concept of fractal dimension can be most easily understood via the so-called box-counting dimension. Let me consider a set M embedded in a d-dimensional space. Let us cover the set with a mesh of d-dimensional cubes with the edge size l and let $N_l(M)$ is a minimal number of the cubes needed for the covering, see Fig. 2.2. The coarse-grained volume $V_l(M) \equiv V_l$ of the set M is then $V_l(M) = l^D N_l(M)$ from which we have the pre-fractal dimension

$$D = \frac{\ln V_l(M)}{\ln l} - \frac{\ln N_l(M)}{\ln l}. \tag{2.10}$$

The box-counting (fractal) dimension of M is then defined as

$$D_{BC} = \lim_{l\to\infty} \frac{\ln V_l(M)}{\ln l} - \frac{\ln N_l(M)}{\ln l} = -\lim_{l\to\infty} \frac{\ln N_l(M)}{\ln l}. \tag{2.11}$$

Here an implicit assumption was made that V_l is finite in the small l limit.

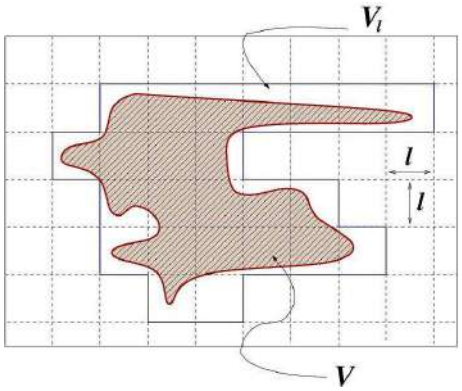
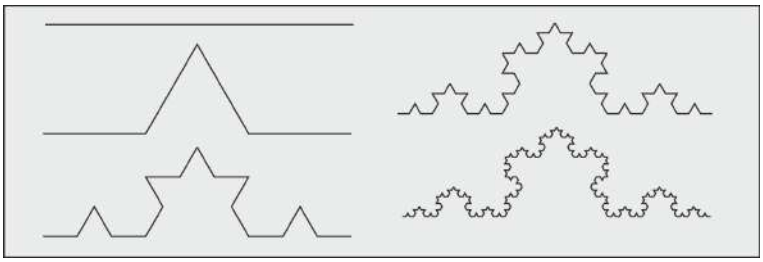


Fig. 2.2: Box counting in 2 dimensional embedding space.

It is quite instructive to illustrate the utility of D_{BC} on a simple example of the so-called *triadic Koch curve*. The latter is defined iteratively in the following way [cf., Fig. 2.3 a)]: in 0th iteration ($n = 0$) I start with a straight line - initiator - with length $r_0 = a$. In the following step ($n = 1$) I raise an equilateral triangle over the middle third of initiator. The result is generator. Its four straight line segments ($N_1 = 4$) have length $r_1 = a/3$ and total length $L[a/3] = 4/3a$. The construction of the Koch curve proceeds by replacing each segment of initiator with generator, i.e., for $n = 2$, $r_2 = (1/3)2a$, $L[(1/3)^2 a] = (4/3)^2 a$ and $N_2 = 16$ [cf., Fig. 2.3 b)], etc. So, when $n = k$, I have $r_k = (1/3)ka$, $L[(1/3)^k a] = (4/3)^k a$ and

a)



b)

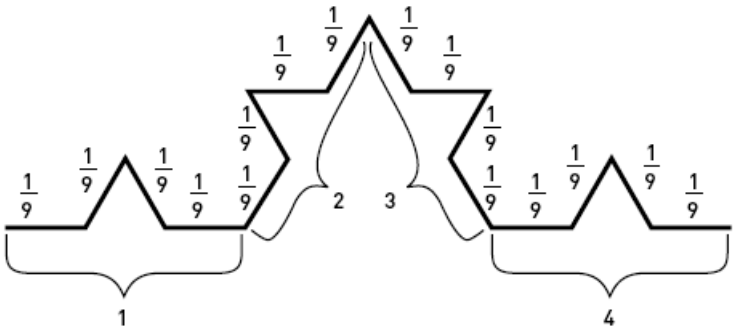


Fig. 2.3: a) Iterative steps in the construction of a triadic Koch curve. Initiator and first four iterations are depicted. b) Details of the 2nd iterative step. It is composed of 4 re-scaled generators and the initiator length r_0 is set to 1.

$N_k = 4k$. Consequently, the box-counting dimension of a triadic Koch curve is

$$D_{BC} = \lim_{r\to 0} \frac{\log N(r)}{\log \frac{1}{r}} = \lim_{n\to\infty} \frac{\log 4^n}{-\log \frac{a}{3^n}} = \frac{\log 4}{\log 3} > 1\,. \quad (2.12)$$

One may often write (e.g., for strictly self.similar fractals), after n iterations $N = N_G^n$ (N_G is the number of pieces of the generator) $r = ar_G^n$ (r_G is the length of the segments of the generator). In such cases the fractal dimension follows from a simple analysis:

$$\lim_{r\to 0} N(r) r^D = \lim_{n\to\infty} \left(N_G r_G^D\right)^n = \text{const.} \;\Rightarrow\; D_{BC} = -\frac{\log N_G}{\log r_G}\,. \quad (2.13)$$

Relation 2.13 allows to recover fairly simply some standard results; e.g., the result for a *triadic Cantor dust* ($r_G = 1/3, N_G = 2$), cf. Fig. 2.4 where $D_{BC} = \log 2/\log 3 < 1$.

I should stress that in most cases of interest the box-counting dimension D_{BC} defined by (2.11) coincides with the Hausdorff–Besicovich fractal

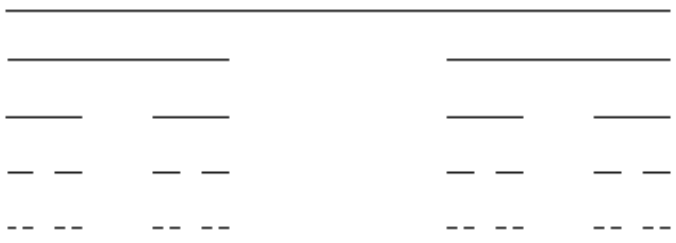


Fig. 2.4: Iterative steps in the construction of a triadic Cantor dust. Initiator and first four iterations are depicted.

dimension D_H used by Mandelbrot [119, 214]. For some exceptional situations, see, e.g., Refs. [80, 81].

Multifractals, on the other hand, are related to the study of a distribution of physical or other quantities on a generic support (be it or not fractal) and thus provide a move from the geometry of sets as such to geometric properties of distributions.

In contrast to ordinary fractals, whose structure is specified by a single scaling exponent — fractal dimension, multifractals have not one but an infinite number of scaling exponents for their description [125] and, in addition, there are *two* qualitatively distinct types of them [116]. To elucidate the idea behind the *first* type, let us suppose that over some support (usually a subset of a metric space) is distributed a probability of a certain phenomenon, be it, e.g., probability of electric charge, magnetic momenta, hydrodynamic vorticity or mass. If I cover the support with a mesh of d-dimensional cubes of size l and denote the integrated probability in the ith cube as p_i, I may define the local scaling exponent α_i via the scaling

$$p_i \ \sim \ l^{\alpha_i} \, . \tag{2.14}$$

At this point I should stress that the existence of such a scaling in one of the defining properties of the multifractal paradigm. The scaling exponent α is known as the Lipshitz–Hölder exponent. Let me denote the number of cubes where p_i has $\alpha_i \in (\alpha, \alpha + d\alpha)$ as $N(\alpha)$, see Fig. 2.5. The *second* type of exponents is determined through the scaling

$$N(\alpha) \ \sim \ l^{-f(\alpha)} \, , \tag{2.15}$$

where the spectral dimension $f(\alpha)$ by its very definition [see Eq. (2.11)] corresponds to a fractal dimension of the region that carries an identical

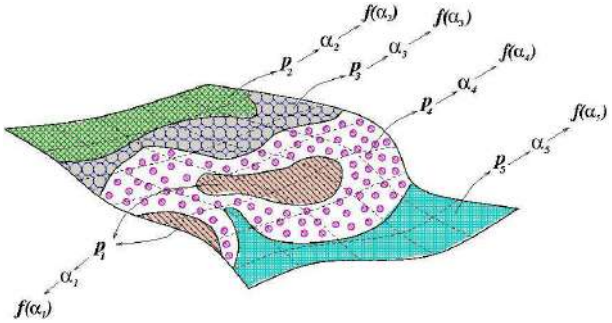


Fig. 2.5: Multifractal: schematic picture

value of p_i. So, in its essence the multifractal is ensemble of intertwined (uni)fractals each with its own fractal dimension $f(\alpha)$. The neologism "multifractal phenomena" describes the fact that different regions of an object have different fractal properties. Multifractal scaling provides a quantitative description of a broad range of heterogeneous phenomena. Paradigmatic example is the so-called *diffusion-limited aggregation* (DLA) that can be observed in many real-world systems ranging from electrodeposition, Hele–Shaw flow, mineral deposits, to dielectric breakdown [14, 334]. Scaling phenomena are so ubiquitous in the nature that the multifractal paradigm is very well suited to describe a large number of complex dynamical systems, see, e.g. Fig. 2.6.

(a) (b) (c) (d) (e)

Fig. 2.6: Some examples of real-world multifractal systems with corresponding support dimensions: soap bubble ($D_H = 2$), particle in turbulent flow ($D_H = 1.38$), Lorentz attractor ($D_H = 2.6$), DLA with cca. 10^3 particles ($D_H = 1.85$), DLA with cca. 10^6 particles ($D_H = 1.71$)

Particularly important applications are also in Statistical analysis of time-series data, namely in finance, climatology or medicine. I will have more

to say about this in Chapter 2.3.2. An extensive and expanding literature now exists on fractals and multifractals, see for instance Refs. [80, 81, 216] and citation therein.

2.2.2 *Rényi entropy and Multifractals*

There is an intimate connection between Rényi entropy and theory of multifractals. It will be seen, that the ensuing relationship is very useful in a wide range of complex structures including many non-linear dynamical systems and their attractors [116, 296]. To this end, I first introduce the so called *Rényi* (or *generalized*) *dimensions* [111, 112, 125, 296, 305]. The Rényi dimension of order q is defined as

$$D_q \;=\; \lim_{\varepsilon\to 0}\frac{1}{(q-1)}\frac{\log\sum_{i=1}^{N(l)} p_i^q}{\log l} \;=\; -\lim_{l\to 0}\frac{\mathcal{I}_q(l)}{\log_2 l}\,, \qquad q\geq 0\,, \quad (2.16)$$

where $N(l) = \sum_i N_l(p_i)$ is the total number of cubes of the edge-length l covering the support of the investigated system. When the superposed physical, biological, etc. phenomena are not considered, then the geometrical structure of the system alone is generally a fractal, see Fig. 2.6. In this case the number of cubes that cover the system scales with l as $N(l) \sim l^{-d}$ with d being the (box-counting) fractal dimension.

As I mentioned in previous section, multifractals are characterized by the (local) probability scaling exponent α and ensuing support fractal dimension $f(\alpha)$. The essence of the multifractal paradigm hinges on the assumption that $f(\alpha)$ is a smooth function of α where α is a continuous variable on $\mathbb{R}_{\geq 0}$. The above premise is often strengthened by assuming that $f(\alpha)$ is differentiable. For example, in chaotic dynamical systems one is often interested in a strange attractor and how often a given region of the attractor is visited in the configuration space [see Fig. 2.6(c)]. In such cases, α is the Lipshitz–Hölder exponent of the visiting PDF (also known as the *distribution of the number of visits*) and $f(\alpha)$ is the Hausdorff fractal dimension of the support of the visiting PDF with the same scaling exponent α. Assumption that $f(\alpha)$ is differentiable grasps very well the observed behavior in strange attractors.

Rather than focus on $f(\alpha)$ and α, it is more convenient for most practical purposes to work with another couple of scaling exponents. These can be obtained from $f(\alpha)$ and α via the Legendre transform. In fact, to keep track with p_i's it is customary in probability theory to define "partition

function" [296]

$$Z(q) \;=\; \sum_i p_i^{\,q} \;=\; \sum_\alpha N(\alpha) p^{\,q}(\alpha)\,. \tag{2.17}$$

Relation (2.17) can be further rewritten as

$$Z(q) \;=\; \int_{\mathbb{R}_{\geq 0}} dN(\alpha) l^{q\alpha} \;=\; \int_{\mathbb{R}_{\geq 0}} d\alpha\; n(\alpha)\, l^{-f(\alpha)+q\alpha}\,. \tag{2.18}$$

Here, $n(\alpha)$ is (weakly l dependent) proportionality function having its origin in relations (2.14) and (2.15). In the small-l limit the asymptotic behavior of the partition function can be evaluated by the steepest descent method [231]. This yields the scaling

$$Z(q) \;\sim\; l^{\tau}\,, \tag{2.19}$$

with

$$\tau(q) \;=\; \inf_{\alpha\in\mathbb{R}_{\geq 0}} [q\alpha - f(\alpha)]\,. \tag{2.20}$$

If the function f is concave and everywhere differentiable then I can rewrite (2.20) as

$$\tau(q) \;=\; q\alpha^* - f(\alpha^*) \quad \text{where} \quad f'(\alpha^*) \;=\; q\,. \tag{2.21}$$

The latter relation defines $\alpha^* = \alpha^*(q) = (f')^{-1}(q)$ since a differentiable concave function has inverse. By differentiating the first identity in (2.21) with respect to q, I obtain

$$\tau'(q) \;=\; \alpha^*\,. \tag{2.22}$$

In following we will simply write α instead of α^*, unless explicitly stated otherwise. This is also a typical convention employed in the literature. Equation (2.20) alongside with its special case (2.21)-(2.22) represents the Legendre–Fenchel transform [272], which for concave functions f boils down to the conventional Legendre transform. Consequently, pairs $\{f(\alpha), \alpha\}$ and $\{\tau(q), q\}$, are convex conjugates that comprise the same mathematical content. Scaling function $\tau(q)$ is called *correlation* or *mass exponent* of the q-th order.

Connection of Rényi entropies with multifractals can be introduced through the concept of generalized dimensions (2.16). In particular, we can observe that

$$D_q(\mathcal{P}) \;=\; -\lim_{l\to 0} \frac{\mathcal{I}_q(\mathcal{P}, l)}{\log_2 l} \;=\; \frac{\tau(q)}{(q-1)}\,. \tag{2.23}$$

In multifractal systems all D_q are necessary to describe uniquely e.g., strange attractors [111, 112, 125]. While the proof of this is based on a rather complicated self-similarity argumentation, by employing the information theory one can show that the assumption of a self-similarity is not really fundamental, and in fact, the validity of the statement extends even beyond the multifractal paradigm [144]. For instance, when the outcome space is discrete then all D_q with $q \in [1, \infty)$ are needed to reconstruct the underlying distribution, and when the outcome space is d-dimensional subset of $\mathbb{R}^d$ then all D_q, $q \in (0, \infty)$ are required to pinpoint uniquely the underlying PDF. The latter examples are nothing but the information theoretic variants of Hausforff's moment problem [144].

In practice, however, only relatively few values of q (typically between 0 and 4) are used. This can be attributed to the fact that only a relatively small number of D_q's have a well defined operational meaning. Notable examples are provided by D_0, D_1 and D_2. In particular, D_0 represents the fractal dimension D_{BC} of the multifractal support. The latter can be seen by employing equations (2.16) and the fact that $Z(q = 0) = N(l) \sim l^{-d}$, where $d = D_{BC}$. Dimension D_1 corresponds to the so-called *curdling dimension*[3] and D_2 coincides with the *correlation dimension* [110]. Note also that by its very definition, D_1 is affiliated with Shannon entropy. This clearly illustrates that SE alone brings only a partial information on multifractal systems.

2.2.3 *Canonical formalism on multifractals*

The connection of Rényi's entropy with multifractal systems can be further deepened in the framework of canonical approach that can be developed in a close analogy with canonical formalism of statistical mechanics. As this approach is thoroughly discussed, e.g., in [144], I will, for shortness's sake, mention only the salient points here.

Let us first consider a multifractal with a density distribution $p(x)$. If we use, as previously, the covering grid with the lattice spacing spacing l then the coarse-grained Shannon's entropy of such a process will be

$$H(\mathcal{P}(l)) = -\sum_k p_k(l) \log_2 p_k(l) \,. \tag{2.24}$$

[3]Verb "curdle" is synonymous with "coagulate". The neologism "curdling dimension" was coined by Mandelbrot, to denote the fractal dimension of the set on which almost all the probability is concentrated, i.e. curdled.

Here the coarse–grained probability distribution $\mathcal{P}(l) = \{p_k(l)\}_{k=1}^{N(l)}$ has the explicit form

$$p_k(l) \;=\; \int_{k-\text{th box}} p(x)\, d^d x\,. \tag{2.25}$$

Let us now observe that from (2.21)-(2.22) and (2.23) we have for $q = 1$

$$\begin{aligned}\alpha(1) = \left.\frac{d\tau(q)}{dq}\right|_{q=1} \;&=\; f(\alpha(1)) \\ &=\; \lim_{l\to 0} \frac{\sum_k p_k(l)\log_2 p_k(l)}{\log_2 l} \;=\; -\lim_{l\to 0}\frac{\mathcal{S}(\mathcal{P}_n(l))}{\log_2 l}\,.\end{aligned} \tag{2.26}$$

In wring this relation I have also used (2.18) and (2.20) to get $0 = \tau(1) = \alpha(1) - f(\alpha(1))$, and (2.19) to write

$$\frac{d\tau(q)}{dq} \;=\; \frac{1}{\log l}\frac{d\log Z(q)}{dq} \;=\; \alpha(q)\,. \tag{2.27}$$

We might also note that $\alpha(1)$ is nothing but $D_1(\mathcal{P})$, i.e., the Hausdorff dimension of the set on which the probability is concentrated — *measure theoretic support*. In fact, the relative probability of the complement set approaches zero when $l \to 0$. This statement is known as Billingsley theorem [37] or curdling [216], and it is a part of a wider family of phenomena know as the *concentration of measure phenomena* [99].

The concavity of $f(\alpha)$ ensures that $\alpha = \alpha(q)$ is monotonically decreasing function of q, and thus $\alpha(q_1) < \alpha(q_2)$ for $q_1 > q_2$. The latter implies that $f(\alpha)$ must terminate at points $\alpha_{min} = \alpha(q = +\infty)$ and $\alpha_{max} = \alpha(q = -\infty)$. Now, because $f'(\alpha) = q$ we have

$$df(\alpha) \;=\; \begin{cases} \leq d\alpha & \text{if } q \leq 1\,, \\ \geq d\alpha & \text{if } q \geq 1\,. \end{cases} \tag{2.28}$$

After integrating the first inequality in (2.28) from $\alpha(q = 1)$ to $\alpha(q)$ and the second inequality from $\alpha(q)$ to $\alpha(q = 1)$, I get that $f(\alpha) \leq \alpha$. This is very important inequality from which we can immediately understand why $D_1 = \alpha(1)$ describes the (curdling) dimension of the measure theoretic support. In fact, we might observe that each unifractal characterized by α_i carries probability $N(\alpha_i)l^{\alpha_i} \sim l^{-f(\alpha_i)+\alpha_i}$. Since $f(\alpha_i) \leq \alpha_i$, we see that in the limit $l \to 0$, the probability of each unifractal tends to zero apart from the situation when $f(\alpha_i) = \alpha_i$ in which case the probability is 1. Due to concavity of $f(\alpha)$ the later will happen only in a single point, namely in the point $\alpha = \alpha(1)$, see Fig. 2.7. Consequently, $\alpha(1)$ describes

not only the Lipshitz–Hölder exponent of the the measure theoretic support but also its fractal dimension $f(\alpha(1))$. In this connection it is interesting to notice that since $\mathcal{I}_0 > \mathcal{I}_1$, I have also $D_0 > D_1$, so the fractal dimension of the multifractal support is bigger than the fractal dimension of measure theoretic support. In a sense, the multifractal support must be *rougher* then the measure theoretic support.

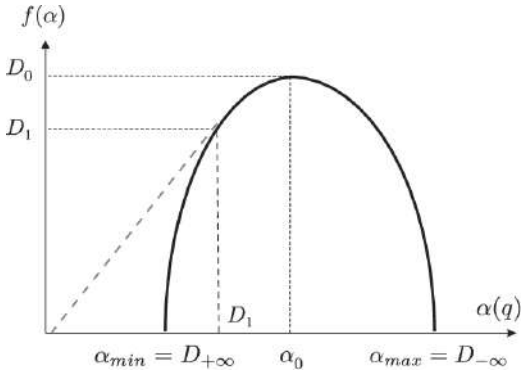


Fig. 2.7: The graph of $f(\alpha)$ versus α is known as the *multifractal spectrum*. Note that the whole curve satisfies the condition $f(\alpha) \leq \alpha$ with equality only in the point $\alpha = \alpha(q = 1) = D_1$. Though all $q \in \mathbb{R}$ can in principle enter into multiftactal analysis, from information theoretic point of view only $q > 0$ are relevant.

For the following considerations it is useful to introduce a one–parametric family of normalized measures $\varrho(q)$ — the so-called zooming or escort distributions [26, 32] [cf. Eq. (2.6)] that is given by

$$\varrho_i(q, l) \;=\; \frac{[p_i(l)]^q}{\sum_j [p_j(l)]^q} \sim l^{f(\alpha_i)} \,. \tag{2.29}$$

As the escort distribution (2.29) alters the scaling of original $\mathcal{P}_n$ from l^α to $l^{f(\alpha)}$, also the measure theoretic support must change. The fractal (curdling) dimension of the new measure theoretic support $\mathcal{M}^{(q)}$ of $\varrho(q)$ is

$$d_h(\mathcal{M}^{(q)}) \;=\; \lim_{l \to 0} \frac{1}{\log_2 l} \sum_k \varrho_k(q, l) \log_2 \varrho_k(q, l) \,. \tag{2.30}$$

Note that the curdling (2.30) mimics the situation that is well known from equilibrium statistical physics. There, in the canonical ensemble formalism one works with (usually infinite) ensemble of identical systems with all possible energy configurations. But only the configurations with E_i close

to $\langle E(T)\rangle$ dominate in the thermodynamic limit. A choice of temperature then shifts the value of $\langle E(T)\rangle$ and hence the support of contributing energy configurations. In statistical physics this behavior is a consequence of the so-called *asymptotic equipartition property* [62].

Let me now define "microcanonical" (unifractal) partition function as

$$Z_{\rm mic}(\alpha_i) \;=\; \left[\sum_{\alpha_k\in[\alpha_i,\alpha_i+d\alpha_i]} 1\right] \;=\; N(\alpha_i)\,. \tag{2.31}$$

This implies, e.g. that

$$\langle\alpha\rangle_{\rm mic} \;=\; \alpha_i \;=\; \sum_{\alpha_k\in[\alpha_i,\alpha_i+d\alpha_i]} \alpha_k/Z_{\rm mic}\,. \tag{2.32}$$

The corresponding microcanonical (Boltzmann-like) entropy is

$$S_{\rm mic}(\alpha_i) \;=\; \log_2 N(\alpha_i) \;=\; \log_2 Z_{\rm mic}(\alpha_i)\,, \tag{2.33}$$

which immediately applies that

$$-\frac{S_{\rm mic}(\alpha_i)}{\log_2 l} \;\sim\; f(\alpha_i) \;\equiv\; \langle f(\alpha)\rangle_{mic}\,. \tag{2.34}$$

Interpreting $E_i = -\alpha_i \log_2 l$ as i-th "energy level", I may define the "inverse temperature" $1/T = \beta/\ln 2$ (note that here $k_B = 1/\ln 2$) as

$$1/T \;=\; \left.\frac{\partial S_{\rm mic}}{\partial E}\right|_{E=E_i} \;=\; f'(\alpha_i) \;=\; q\,. \tag{2.35}$$

On the other hand, I can define the "canonical" partition function as

$$Z_{\rm can}(q) \;=\; \sum_i p_i(l)^q \;=\; \sum_i e^{-\beta E_i}\,, \tag{2.36}$$

where $\beta = q\ln 2$ and $E_i = -\log_2 p_i(l)$. The corresponding mean values are

$$\begin{aligned} f(q) \;&\equiv\; \langle f(\alpha)\rangle_{\rm can} \;=\; \sum_i \frac{f(\alpha_i)}{Z_{\rm can}}\, e^{-\beta E_i} \;\sim\; \frac{\sum_i \varrho_i(q,l)\log_2\varrho_i(q,l)}{\log_2 l}\,, \\ \alpha(q) \;&\equiv\; \langle\alpha\rangle_{\rm can} \;=\; \sum_i \frac{\alpha_i}{Z_{\rm can}}\, e^{-\beta E_i} \;\sim\; \frac{\sum_i \varrho_i(q,l)\log_2 p_i(l)}{\log_2 l}\,. \end{aligned} \tag{2.37}$$

Let us note that the fractal dimension of the measure theoretic support $d_h(\mathcal{M}^{(q)})$ is simply $f(q)$. The previous results are gathered in Tab. 2.1:

microcanonical ensemble - unifractals	**canonical ensemble - multifractals**
$Z_{\rm mic}$; $S_{\rm mic} = \log_2 Z_{\rm mic}$	$Z_{\rm can}$; $S_{\rm can} = \log_2 Z_{\rm can} - q\langle\alpha\rangle_{\rm can} \log_2 l$
$\langle\alpha\rangle_{\rm mic} = \alpha_i = \sum_k \alpha_k / Z_{\rm mic}$	$\langle\alpha\rangle_{\rm can} = \sum_k \alpha_k \; p_k^q / Z_{\rm can}$
$\langle f(\alpha)\rangle_{\rm mic} = f(\alpha_i) = -S_{\rm mic}/\log_2 l$	$\langle f(\alpha)\rangle_{\rm can} = -S_{\rm can}/\log_2 l$
$q = \partial S_{\rm mic}/\partial E\vert_{E=E_i}$	$q = \partial S_{\rm can}/\partial\langle E\rangle_{\rm can}$
$1/T = q \;\; (k_B \cong 1/\ln 2)$	$1/T = q \;\; (k_B \cong 1/\ln 2)$
$E_i = -\log_2 p_i = -\alpha_i \log_2 l$	$\langle E\rangle_{\rm can} = -\langle\alpha\rangle_{\rm can} \log_2 l$
$\langle f(\alpha)\rangle_{\rm mic} = q\langle\alpha\rangle_{\rm mic} - \tau$	$\langle f(\alpha)\rangle_{\rm can} = q\langle\alpha\rangle_{\rm can} - \tau$

Tab. 2.1. Parallelism between microcanonical and canonical description in multifractals.

It is worth of noting that the last lines are basically the first lines in disguise. In addition, by defining the "Helmholtz free energy" $F = -q \log_2 Z_{\rm can} = -\beta \ln Z_{\rm can}$, I get from the first/last identity the usual thermodynamical relation

$$F \;=\; \langle E\rangle_{\rm can} \;-\; S_{\rm can} T \,. \tag{2.38}$$

By, looking at fluctuations of α in the "canonical" ensemble one can easily establish an equivalence between microcanonical and canonical description of multifractals. Let us first observe that

$$\begin{aligned} \partial^2(\log_2 Z_{\rm can})/\partial q^2 \;&=\; \langle E^2\rangle_{\rm can} - \langle E\rangle_{\rm can}^2 \\ &=\; \left(\langle\alpha^2\rangle_{\rm can} - \langle\alpha\rangle_{\rm can}^2\right) (\log_2 l)^2 \,. \end{aligned} \tag{2.39}$$

Despite its appearance, the true scaling of $\mathrm{Var}(E)$ is $\log_2 l$ and not $(\log_2 l)^2$. Indeed, recalling Eq. (2.19) I can write

$$\partial^2(\tau \log_2 l)/\partial q^2 \;=\; (\partial\alpha/\partial q) \log_2 l \;\propto\; \log_2 l \,. \tag{2.40}$$

So, the relative standard deviation of "energy" reads

$$\frac{\sqrt{\langle E^2\rangle_{\rm can} - \langle E\rangle_{\rm can}^2}}{\langle E\rangle_{\rm can}} \;=\; \frac{\sqrt{\langle\alpha^2\rangle_{\rm can} - \langle\alpha\rangle_{\rm can}^2}}{\langle\alpha\rangle_{\rm can}} \;\propto\; \frac{1}{\sqrt{-\log_2 l}} \;\to\; 0 \,. \tag{2.41}$$

Consequently, in the small-l limit (i.e., exact multifractal limit) the α-fluctuations become negligible and all relevant α_i are sharply peaked (curdled) around $\langle\alpha\rangle_{\rm can}$, provided q is a solution of the equation $\alpha_i = \tau'(q)$, see Fig 2.8.

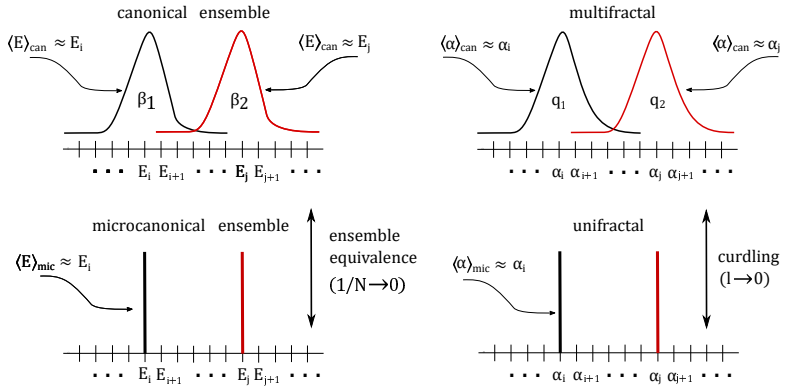


Fig. 2.8: Comparison of the microcanonical and canonical description in equilibrium statistical physics and in multifractal theory.

The foregoing implies that in the "thermodynamic" limit ($l \to 0$)

$$f(\alpha_i) \;=\; \langle f(\alpha)\rangle_{\rm mic} \;=\; \langle f(\alpha)\rangle_{\rm can} \;=\; f(q(\langle\alpha\rangle_{\rm can}))\,, \qquad (2.42)$$

since the RHS's of the last equations in Tab. 2.1 are identical. This in turn implies (see, 3rd line in Tab. 2.1) that the microcanonical and canonical entropies on multifractal coincide. Hence

$$S_{\rm mic}(\alpha_i) \;\sim\; -\sum_k \varrho_k(q,l)\log_2\varrho_k(q,l) \;\equiv\; \mathcal{S}(\varrho)|_{f(q)}\,.$$

The subscript $f(q)$ emphasizes that the Shannon entropy $\mathcal{S}(\varrho)$ is basically the entropy of an unifractal specified by the fractal dimension $f(q)$ with $q = f'(\langle\alpha\rangle_{\rm can}) \sim f'(\alpha_i)$. Legendre transform then implies that

$$S_{\rm mic}(\alpha_i) \;\sim\; -q\langle\alpha\rangle_{\rm can}\log_2 l \;+\; (1-q)\mathcal{I}_q(\mathcal{P}(l))\,. \qquad (2.43)$$

In the continuum limit when $l \to 0$, one can employ the renormalization prescription which leads to finite entropies — so-called continuous (or differential) entropies [144]. With this one gets

$$\mathcal{I}_q^r \;=\; S_{\rm mic}^r(\alpha_i)\,, \qquad (2.44)$$

where superscript r indicates renormalized quantities. So, by changing the q parameter in Rényie's entropy one can "skim over" all renormalized unifractal entropies. Rényi's entropy thus provides a unified information measure that keeps track of all respective unifractal entropies.

The passage from multifractals to single-dimensional statistical systems is done by assuming that the α-interval is infinitesimally narrow and that the corresponding PDF is smooth. In such a case both α and $f(\alpha)$ collapse to $\alpha = f(\alpha) \equiv D$ and $q = f'(\alpha) = 1$. For instance, for a statistical system with a smooth measure and the support space $\mathbb{R}^d$ one can see that Eq. (2.44) constitutes a trivial identity. This is the primary reason why Shannon's entropy plays such a predominant role in physics of single-dimensional sets. Further discussion of the relation (2.44) can be found in [144].

2.2.4 *Rényi's entropy and Fisher's information*

There exits an interesting connection between Riemaniann geometry on statistical parameter spaces (so-called information geometry) and Rényi entropies.

To see this let us consider a family of PDF's characterized by a vector parameter θ

$$\mathcal{F}_\theta \;=\; \{p(x,\theta); x \in M; \theta \in \mathcal{M}, \text{a manifold in} \mathbb{R}^n\}\;. \tag{2.45}$$

Let me further assume that $p(x,\theta) \in \mathcal{C}^2$. The Gibbs PDF's (with θ_i being the inverse temperature β, the external field H, etc.) represent particular examples of (2.45).

To construct a metric on $\mathcal{M}$ which reflects the statistical properties of the family (2.45), Rao a co-workers [48, 49] proposed to adopt various measures of dissimilarity between two probability densities, and then use them to derive the metric. Important class of dissimilarity measures are information-theoretic measures and, in particular, the gain of information when a density $p(x,\phi)$ is replaced with a density $p(x,\theta)$. In Shannonian information theory the gain of information is represented be the Kullback–Leibler divergence (also called relative entropy), i.e.

$$D(p_\theta||g_\phi) \;\equiv\; (\theta||\phi) \;=\; \int_M dx\, p(x,\theta) \log_2 \left(\frac{p(x,\theta)}{p(x,\phi)}\right). \tag{2.46}$$

In the case of Rényi's entropy the gain of information (or relative entropy of order q) has the forms [144, 264]

$$\mathcal{I}_q(\theta||\phi) \;=\; \frac{1}{q-1} \log_2 \int_M dx\; \frac{p(x,\theta)^q}{p(x,\phi)^{q-1}}\,. \tag{2.47}$$

At this point I should stress that the relative entropy cannot be understand as the "distance" between probability distributions, because it does not satisfy basic properties of a distance function (i.e., metric). It does not obey the triangle inequality, and in general $D(p_\theta||g_\phi)$ does not equal $D(p_\phi||g_\theta)$ (same for Rényi's relative entropy). However, its infinitesimal form has correct distance function properties. Indeed, let me look at the leading order of dissimilarity between $p(x,\theta)$ and $p(x,(\theta+d\theta))$, namely

$$\mathcal{I}_q(\theta||\theta+d\theta) \;=\; \frac{1}{2!}\sum_{i,j} g_{ij}(\theta)\; d\theta_i d\theta_j \;+\; \ldots\,. \tag{2.48}$$

Note that because $\mathcal{I}_q(\theta||\phi)$ is minimal at $\theta=\phi$, the linear term in (2.48) vanishes. So we have

$$\begin{aligned} g_{ij}(\theta) &= \left[\frac{\partial^2}{\partial\phi_i\partial\phi_j}\mathcal{I}_q(\theta||\phi)\right]_{\theta=\phi} \\ &= \frac{q}{2\log 2}\left(\int_M dx\; p(x,\theta)\frac{\partial\log p(x,\theta)}{\partial\theta_i}\,\frac{\partial\ln p(x,\theta)}{\partial\theta_j}\right) \\ &= \frac{q}{2\log 2}\; F_{ij}(p(x,\theta)\,. \end{aligned} \tag{2.49}$$

Since the Hessian F_{ij} is a symmetric matrix, Eq. (2.49) represents a metric tensor, which in the information geometry is known as the Fisher information matrix (or Fisher–Rao metric) [48, 49, 259]. Fisher matrix is the only Riemaniann metric which is invariant under transformation of variables as well as reparametrization [259]. In addition, the diagonal elements of Fisher's information matrix represent the amount of information on θ_i in an element of a sample chosen from a population of density functions $p(x,\theta)$. Due to its relation with Cramér–Rao inequality Fisher information matrix plays a crucial rôle in parametric estimation [93]. Let me also stress that the latter is used in quantum mechanics to formulate information uncertainty relations [93, 146].

The information geometry is presently a fast growing field in statistical physics and statistical inference theory with a number or important results. An interested reader can consult e.g., Ref. [5] for further details.

2.3 Financial markets and econophysics - I

Particularly important application of the Rényi's entropy paradigm is in *econophysics*. Econophysics is an emerging interdisciplinary field that uses

a deep analogy of economical and financial processes with physical phenomena (treated in many body physics, physical kinetics, statistical mechanics and even quantum theory) in order to analyze, e.g., the growth or death of the company, profit rate and company size optimization, as well as the behavior, growth and failure of markets [45, 222].

2.3.1 *Multifractal financial markets*

The evolution of many complex systems in natural, economical, medical and biological sciences is usually presented in the form of time data sequences. A global massification of computers together with their improved ability to collect and process large data-sets has brought about the need for novel analyzing methods. Particularly in the connection with financial time series there has been rapid development of techniques for measuring and managing the fractal and multifractal scaling behavior observed in empirical high-frequency data sequences. A non-trivial scaling behavior in a time data-set represents a key signature of a multi-time scale cooperative behavior in much the same way as a non-trivial scaling behavior in second-order phase transitions reflects the underlying long-range (or multi-scale) cooperative interactions. The usefulness of the scaling approach is manifested, for instance, in quantifying critical or close-to-critical scaling, which typically signalizes the onset of financial crises, including stock market crashes, currency crises or sovereign defaults [179]. A multifractal scaling, in particular, is instrumental in identifying the relevant scales that are involved in both temporal and inter-asset correlations [151]. In passing, I should stress that aside from financial data sequences, similar (multi)fractal scaling patterns are also routinely observed (and analyzed) in time data-sets found, for instance, in heart rate dynamics [252, 326], DNA sequences [222, 251] or long-time weather records [304].

One can establish connection between financial time series and multifractals through generalized dimensions (2.16) and (2.23). The point is that D_q can be obtained from high-quality time series data if care is taken. This is done by employing some convenient estimators of Rényi entropy. For instance, I can partition the configuration space into disjoint cubes or balls where sides or radiuses have size l and introduce the coarse-grained distribution [cf. Eq. (2.25)]

$$p_l(x_k) \;=\; \int_{U_l(x_k)} d^d x p(x) \;=\; \int_{U_l(x_k)} d\mu(x)\,, \tag{2.50}$$

where $d\mu(x)$ is ensuing probability measure and $U_l(x_k)$ is a ball of radius l

centered at x_k. With this I get the coarse-grained RE

$$\begin{aligned}\mathcal{I}_q(\mathcal{P},l) &= \frac{1}{1-q}\log_2\sum_k [p_l(x_k)]^q \\ &= \frac{1}{1-q}\log_2\sum_k \int_{U_l(x_k)} d\mu(x)[p_l(x_k)]^{q-1} \\ &\cong \frac{1}{1-q}\log_2\int_{\mathbb{R}^d} d\mu(x)\left[\int_{\mathbb{R}^d} d\mu(y)\,\theta(l-\|x-y\|)\right]^{q-1}. \quad (2.51)\end{aligned}$$

Here $\theta(x)$ is the Heaviside step function, i.e., $\theta(x) = 0$ for $x \leq 0$ and $\theta(x) = 1$ for $x > 0$. The last line holds in the small-l limit where I can assume that $p_l(x_k) \cong p_l(x)$ for any x within the $U_l(x_k)$ ball.

Formula (2.51) allows to define an estimator of the RE from a finite sequence of N points as

$$\hat{\mathcal{I}}_q(l) = \frac{1}{(1-q)}\log_2\left\{\frac{1}{N(N-1)^{q-1}}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\theta(l-\|x_i-y_j\|)\right]^{q-1}\right\}, \quad (2.52)$$

where an implicit assumption was made that each point contributes with equal weight that is inverse proportional to the number of points in the considered sums.

Estimator (2.52) is of practical value in cases when $q > 1$. For $q \leq 1$ it runs into severe problems whenever inside of some ball there are no points. This can be rectified by another estimators such as, e.g., fixed-mass algorithm [11] (where the radius l of balls $U_l(x_k)$ is not fixed but can vary so that each ball contains an equal number of points), Kozachenko–Leonenko estimator [183] or various Leonenko *et al.*, entropic estimators [194].

2.3.2 *Causality and Rényi's information transfer between financial time series*

With a proper estimator of the RE one can attack number of important issues in data analysis. Particularly pertinent application concerns the *causality* between two or more time series.

The first general definition of causality, which could be quantified and measured computationally was given by Wiener in 1956, namely "*... For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one...*" [331].

2.3.2.1 *Granger causality ad other causal measures*

The introduction of the concept of causality into the experimental practice, namely into analyses of data observed in consecutive time instants (i.e., time series) is due to Nobel prize winner (economy, 2003) C.W.J. Granger.

Definition 2.1 (Granger causality). *Process Y_t Granger causes another process X_t if future values of X_t can be better predicted using the past values of Y_t and X_t rather than only past values of X_t.*

The standard test of Granger causality was developed by Granger himself [109] and it is based on a linear regression model, namely

$$X_t \;=\; a_0 \;+\; \sum_{\ell=1}^{n} a_{1\ell} X_{t-\ell} \;+\; \sum_{\ell=1}^{n} a_{2\ell} Y_{t-\ell} \;+\; e_t\,, \tag{2.53}$$

where n represents the maximum number of lagged observations included in the model (the model order), t is a discrete time with possible discrete values $n+1,\ldots,N$ and e_t is uncorrelated random variable (random error) with zero mean and variance σ^2. The *null hypothesis* that Y_t does not cause X_t (in the sense of Granger) is accepted if and only if $a_{2\ell} = 0$ for $\ell = 1,\ldots,n$, which reduces (2.53) to

$$X_t \;=\; a_0 \;+\; \sum_{\ell=1}^{n} a_{1\ell} X_{t-\ell} \;+\; \tilde{e}_t\,. \tag{2.54}$$

Here the noise term $\tilde{e}_t$ might be generically different (if $\sigma_1 \neq \sigma_2$). This null hypothesis can be tested by various means. Particularly popular statistical tests is F-test of the null hypothesis that $a_{2\ell} = 0$ for $\ell = 1,\ldots,n$. In this case one defines the F-statistic (the Grange–Sargent statistic) [129] as

$$F \;=\; \frac{(RSS_1 - RSS_2)}{df_1 - df_2} \Big/ \frac{RSS_2}{df_2} \;=\; \frac{(RSS_1 - RSS_2)/n}{RSS_2/(N-3n)}\,, \tag{2.55}$$

where RSS_1 and RSS_2 are sums of squares of residual errors after (2.54) and (2.53) are fitted to the same data set. The degrees of freedom — denoted as df_1 and df_2 — are associated with the model (2.54) and (2.53), respectively. In particular, in the null-hypothesis model (2.54) $df_1 = N - 2n$. Indeed, we start with $N - n$ data points (degrees of freedom) but out of those, n data points are needed for estimation of parameters $a_{11},\ldots,a_{1n}$ that enter in RSS_1. Parameter a_0 can be find from the consistency requirement that the mean error is zero[4]. Similarly, for (2.53) we get $df_2 = N - 3n$.

[4]Note that, e.g.,

$$RSS_1 \;=\; \sum_{t=n+1}^{N} (x_t - \bar{X}_t)^2\,, \tag{2.56}$$

Note that F itself is a random variable, which under assumption that the residual errors from the two models are i) independent and ii) normally distributed (which incidentally happen to be requirements of *ordinary least squares regression*) has the Fisher F-distribution $F(n, N-3n)$.

Aforementioned linear framework for measuring and testing causality has been widely applied in a number of fields. In finance, one typically uses Granger's linear regression model to study the internal cross-correlations between various market activities. The correlation functions have, however, at least two limitations: First, they measure only linear relations, although it is clear that linear models do not faithfully reflect real market interactions. Second, all they determine is whether two time series (e.g., two stock-index series) have correlated movement. They, however, do not indicate which series affects which, or in other words, they do not provide any *directional information* about cause and effect.

Although there is an extensive literature on causality modeling that goes beyond the linear regression model, e.g., applying and combining mathematical logic, graph theory, Markov models, Bayesian probability, etc. (for extensive review see, e.g., [250]), I will focus here mostly on the information-theoretic approaches which understand causality as a phenomenon which can be not only detected or measured but also quantified.

Particularly important quantifier of the information flow between two time series is the so-called *transfer entropy* (TE). In order to illustrate what is involved I will first start with the concept of Shannonian transfer entropy and then generalize it to the RE setting.

2.3.2.2 *Shannon's transfer entropy*

Concept of TE was introduced by Schreiber in his 2000 seminal paper [280] and independently under the name *conditional mutual information* by Paluš *et al.* in Ref. [248]. According to these, TE represents a measure of a directional (Shannonian) information flow defined by means of *Kullback–Leibler divergence* on conditional transition probabilities of two finite-order Markov processes X_t and Y_t. An advantage of information theoretic measures, as compared to standard Granger causality, is that they are sensitive to nonlinear signal properties as they do not rely on linear regression models. A limitation of transfer entropies, in comparison to Granger causality, is that they

where x_t are empirical data and $\bar{X}_t = a_0 + \sum_{\ell=1}^{n} a_{1\ell} X_{t-\ell}$ are fitted data. Actual parameters in the fitted data are obtained from the method of minimal squared errors under additional constraint that the mean error $\sum_t^N (x_t - \bar{X}_t) = 0$.

are by their very formulation restricted to bivariate situations. In addition, information theoretic measures often require substantially more data than regression methods. For a comparison of transfer entropy with other causal measures, including various implementations of Granger causality, see e.g., Ref. [205]. It can be also shown that for Gaussian variables, Granger causality and transfer entropy are entirely equivalent [17].

Let us now consider two time sequences (e.g., two stock market time series) described by stochastic random variables X_t and Y_t. Let us assume further that the time steps (e.g., data ticks) are discrete with the time step τ and with $t_n = t_0 + n\tau$ where t_0 is some reference time. Shannon's TE $T_{Y\to X}(m,l)$ is defined as [248, 280]

$$\begin{aligned} T_{Y\to X}(m,l) &= H(X_{t_{m+1}}|X_{t_1},\ldots,X_{t_m}) \\ &\quad - H(X_{t_{m+1}}|X_{t_1},\ldots,X_{t_m},Y_{t_{m-l+1}},\ldots,Y_{t_m}) \\ &= I(X_{t_{m+1}};X_{t_1},\ldots,X_{t_m},Y_{t_{m-l+1}},\ldots,Y_{t_m}) \\ &\quad - I(X_{t_{m+1}};X_{t_1},\ldots,X_{t_m})\,. \end{aligned} \tag{2.57}$$

Here $I(;)$ denotes mutual information, cf. Eq. (1.44). In (2.57) I have assumed that X_t and Y_t are Markovian processes of order m and l, respectively. The actual values m and l (characterizing a memory effects) typically arise from numerical tests.

Let us notice that the last two lines of (2.57) can be rephrased as follows: $T_{Y\to X}(m,l)$ represents *gain of information* about $X_{t_{m+1}}$ caused by the *whole history* of X and Y up to time t_m minus *gain of information* about $X_{t_{m+1}}$ caused by the *whole history* of X up to time t_m. This is, however, nothing but *gain of information* about $X_{t_{m+1}}$ caused *purely by the whole history* of Y up to time t_m. With the definition of conditional mutual entropy one can recognize that the last two lines represent the conditional mutual entropy [cf. Eq. (1.44)], namely

$$T_{Y\to X}(m,l) \;=\; I(X_{t_{m+1}};Y_{t_{m-l+1}},\ldots,Y_{t_m}|X_{t_1},\ldots,X_{t_m})\,. \tag{2.58}$$

With the help of the first equality in (2.57), I can explicitly rewrite TE as

$$\begin{aligned} T_{Y\to X}(m,l) &= \sum p(x_{t_1},\ldots,x_{t_{m+1}},y_{t_{m-l+1}},\ldots y_{t_m}) \\ &\quad \times\ \log_2 \frac{p(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m},y_{t_{m-l+1}},\ldots,y_{t_m})}{p(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m})} \\ &\equiv \sum_{x\in X,y\in Y} p(x_{m+1},x_m^{(m)},y_m^{(l)}) \log_2 \frac{p(x_{m+1}|x_m^{(m)},y_m^{(l)})}{p(x_{m+1}|x_m^{(m)})}\,, \end{aligned} \tag{2.59}$$

where x_{t_i} and y_{t_j} represent the discrete-valued states (alphabet) of random variables X_{t_i} and Y_{t_j}, respectively. On the last line I used a concise notation $x_{m+1} = x_{t_{m+1}}$ and $y_m^{(l)} \equiv (y_{t_m}, \ldots, y_{t_{m-l+1}})$ (similarly for $x_m^{(m)}$) with m and l denoting corresponding Markovian orders. The sum runs over all possible states of the random variables $X_{t_1}, \ldots, X_{t_{m+1}}, Y_{t_{m-l+1}}, \ldots, Y_{t_m}$.

A few observations related to $T_{Y\to X}$ are in order:

i) From the first line of Eq. (2.57) we see that $T_{Y\to X} \geq 0$. This is a consequence of Gibbs inequality (see Appendix D.1). This might be restated as saying that any extra knowledge in conditional entropy lessens our ignorance.

ii) Since Gibbs' inequality is saturated if and only if

$$\frac{p(x_{t_{m+1}}|x_{t_1}, \ldots, x_{t_m}, y_{t_{m-l+1}}, \ldots, y_{t_m})}{p(x_{t_{m+1}}|x_{t_1}, \ldots, x_{t_m})} \;=\; 1\,, \tag{2.60}$$

for all states involved, TE is zero if and only if the history of Y up to time t_m has no influence on the value of $X_{t_{m+1}}$ or, in other words, when *there is no information flow* from Y to X; i.e., the Y and X time series are independent processes.

iii) $T_{Y\to X}$ is clearly explicitly non-symmetric (directional) since it measures the degree of dependence of X on Y and not vice versa.

The definition (2.57) allows to interpret the transfer entropy $T_{Y\to X}$ as a rating factor which quantifies a gain/loss in the risk concerning the behavior of X at some future time t_{m+1} after we have taken into account the historical values of a time series Y until t_m, see Ref. [151].

2.3.2.3 *Rényi's transfer entropy*

Rényi's transfer entropy can be defined in much the same way as its Shannon's counterpart [151, 157]. In particular, we can utilize the concept of mutual information of order q and define the Rényi's transfer entropy (RTE) of order q as

$$\begin{aligned} T^{(R)}_{q;Y\to X}(m,l) \;&=\; \mathcal{I}_q(X_{t_{m+1}}|X_{t_1}, \ldots, X_{t_m}) \\ &-\; \mathcal{I}_q(X_{t_{m+1}}|X_{t_1}, \ldots, X_{t_m}, Y_{t_1}, \ldots, Y_{t_1}) \\ &=\; \mathcal{I}_q(X_{t_{m+1}}; Y_{t_1}, \ldots, Y_{t_1}|X_{t_1}, \ldots, X_{t_m})\,. \end{aligned} \tag{2.61}$$

With the help of (1.41) and (1.44) this can be written in an explicit form as

$$
\begin{aligned}
&T^{(R)}_{q;Y\to X}(m,l)\\
&\quad= \frac{1}{1-q}\log_2 \frac{\sum \varrho_q(x_{t_1},\ldots,x_{t_m})p^q(y_{t_{m-l+1}},\ldots,y_{t_m}|x_{t_1},\ldots,x_{t_m})}{\sum \varrho_q(x_{t_1},\ldots,x_{t_{m+1}})p^q(y_{t_{m-l+1}},\ldots,y_{t_m}|x_{t_1},\ldots,x_{t_{m+1}})}\\
&\quad= \frac{1}{1-q}\log_2 \frac{\sum \varrho_q(x_{t_1},\ldots,x_{t_m})p^q(x_{t_{m+1}}|x_{t_1},\ldots,x_{t_m})}{\sum \varrho_q(x_{t_1},\ldots,y_{t_m})p^q(x_{t_{m+1}}|x_{t_1},\ldots,y_{t_m})}\\
&\quad\equiv \frac{1}{1-q}\log_2 \frac{\sum \varrho_q(x^{(m)}_m)p^q(x_{m+1}|x^{(m)}_m)}{\sum \varrho_q(x^{(m)}_m,y^{(l)}_m)p^q(x_{m+1}|x^{(m)}_m,y^{(l)}_m)}\,. \qquad (2.62)
\end{aligned}
$$

Here, ϱ_q is the escort distribution. It can be easily seen that in the limit $q\to 1$ we regain Shanonn's transfer entropy (2.59).

A few comments related to $T^{(R)}_{q;Y\to X}(m,l)$ are now in order:

i) When the history of Y up to time t_m has no influence on the value of the random variable $X_{t_{m+1}}$, then from (2.62) (second line) it follows that $T^{(R)}_{q;Y\to X}(m,l)=0$, which implies that no information flows from Y to X, as it should be expected.

ii) Opposite implication does not hold (unlike in Shannon's case), namely $T^{(R)}_{q;Y\to X}=0$ does not imply independence of X and Y processes. This is because $T^{(R)}_{q;Y\to X}$ can also be *negative*. The reason for this is not difficult to understand. Rényi's entropy works with rescaled distributions, and so it allows to address information flow between different parts of underlying distributions in bivariate time series. For instance for, $q<1$ the negativity of $T^{R}_{q,Y\to X}$ simply means that the knowledge of historical values of both X and Y flattens the *tail part* of the anticipated distribution function for the price value $X_{t_{m+1}}$ more than historical values of X alone would do. In other words, extra knowledge of historical values of Y shows that there is a greater risk in the next time step of X than one would expected by only knowing the historical data of X. In this sense $T^{R}_{q;Y\to X}$ represents a *rating factor* which quantifies a gain or loss in the risk concerning the behavior of X at the future time $X_{t_{m+1}}$ after the historical values of Y until Y_{t_m} were accounted for [151]. In other words, Rényi's TE provides more detailed information concerning the excess (or lack) of information in various parts of the underlying distribution resulting from updating the distribution on the condition

that a second time series is known. This can be particularly relevant in the context of financial time series where the knowledge of tail-part (or "black swan") events such as spikes or sudden jumps bears direct implications, e.g., in various risk-reducing formulas in portfolio theory.

iii) $T^{(R)}_{q;Y\to X}$ is again explicitly directional since it measures the degree of dependence of X on Y and not the other way around, though in this case we should indicate, e.g. by an arrow, whether the original risk rate about $X_{t_{m+1}}$ was increased or reduced by observing the historical values of Y.

iv) In order to view $T^{(R)}_{q;Y\to X}$ as a genuine transfer entropy, one should really include in it the whole history of Y and X up to time t_m (i.e., all historical data that may be responsible for cross-correlations with $X_{t_{m+1}}$). The history is finite only if X or/and Y processes are finite-order Markovian processes. In particular, if X is a Markov process of order $m+1$ and Y is of order l, then $T^{(R)}_{q;Y\to X}(m,l)$ is a true transfer entropy. Unfortunately, most dynamical systems cannot be mapped on Markovian processes with finite-time memory. For such systems one should consider limits $m\to\infty$ and $l\to\infty$. However, in practice the finite size of any real data set hinders this limiting procedure. In order to avoid unwanted finite-size effects, one might define the *effective transfer entropy* as [151, 223]

$$T^{\text{eff}}_{q;Y\to X}(m,l) \;\equiv\; T_{q;Y\to X}(m,l) \;-\; T_{q;Y_{\text{schuffled}}\to X}(m,l)\,. \quad (2.63)$$

Effective RTE is simply a difference between two RTEs, where the second one is computed on the *shuffled* Y series. Here the shuffling is done in terms of the *surrogate data technique* [167]. In its essence, a surrogate data series has the same mean, the same variance, the same autocorrelation function, and therefore the same power spectrum as the original series, but phase relations are destroyed. Consequently, all the potential correlations between X an Y are removed, which implies that $T^{R}_{q;Y_{\text{shuffled}}\to X}$ should be zero. In practice, this is typically not the case, despite the fact that there is no obvious structure in the data. The non-zero value of $T^{R}_{q;Y_{\text{shuffled}}\to X}$ must then be a consequence of the finite data set. Definition (2.63) then simply ensures that pseudo-effects caused by finite values of m and l are subtracted.

So, Rényi's TE defined in (2.62) has many specific properties that are desirable for the quantification of an information flow between two interrelated

stochastic systems. In particular, RTE can serve as an efficient rating factor, which quantifies a gain or loss in the risk that is inherent in the passage from X_{t_m} to $X_{t_{m+1}}$ when a new information, namely historical values of a time series Y until time t_m, is taken into account. This gain/loss is parameterized by a single parameter, the Rényi q parameter, which serves as a "zooming index" that emphasizes or de-emphasizes different sectors of the underlying empirical PDF. In this way one can scan various sectors oft the price distribution and analyze associated information flows. In particular, the fact that one may separately scrutinize information fluxes between tails or central-peak parts of asset price distributions simply by setting $q < 1$ or $q > 1$, respectively, can be employed, for example, by financial institutions to quickly analyze the global (across-the-border) information flows and use them to redistribute their risk. For instance, if an American investor observes that a certain market, say the S&P500, is going down and he/she knows that the corresponding NASDAQ effective RTE for $q < 1$ is low, then he/she does not need to relocate the portfolio containing related assets rapidly, because the influence is in this case slow. Slow portfolio relocation is generally preferable, because fast relocations are always burdened with excessive transaction costs. Let us stress that this type of conduct could not be deduced from Shannon's transfer entropy alone. In fact, the effective Shannon's TE might suggests a fast (and thus expensive) portfolio relocation as a best strategy [151].

2.4 Information Theoretic Uncertainty Relations in Quantum Mechanics

Quite unexpected but very fruitful application of Rényi entropies is in the field of quantum-mechanical uncertainty relations (UR).

2.4.1 *Uncertainty Relations in Quantum Mechanics — a bit of history*

Quantum-mechanical uncertainty relations place fundamental limits on the accuracy with which one is able to measure the values of different physical quantities. This has profound implications not only on the microscopic but also on the macroscopic level of physical systems. The archetypal uncertainty relation formulated by Heisenberg in 1927 describes a trade-off between the error of a measurement to know the value of one observable

and the disturbance caused on another complementary observable so that their product should be no less than a limit set by $\hbar$. Since Heisenberg's intuitive, physically motivated deduction of the error-disturbance uncertainty relations [124], a number of methodologies trying to improve or supersede this result have been proposed. In fact, over the years it have became steadily clear that the intuitiveness of Heisenberg's version cannot substitute mathematical rigor and it came as no surprise that the violation of the Heisenberg's original relation was recently reported a number of experimental groups, e.g., most recently by the Vienna group in neutron spin measurements [79]. At present it is Ozawa's universally valid error-disturbance relation [244, 245] that represents a viable alternative to Heisenberg's error-disturbance relation.

Yet, already at the end of 1920s Kennard and independently Robertson and Schrödinger reformulated the original Heisenberg (single experiment, simultaneous measurement, error-disturbance) uncertainty principle in terms of a statistical ensemble of identically prepared experiments [166, 271, 281]. Among other things, this provided a rigorous meaning to Heisenberg's imprecisions ("Ungenauigkeiten") δx and δp as standard deviations in position and momenta, respectively, and entirely avoided the troublesome concept of simultaneous measurement. The Robertson–Schrödinger approach has proven to be sufficiently versatile in accommodating other complementary observables apart from x and p, such as components of angular momenta, or energy and time. Because in the above cases the variance is taken as a "measure of uncertainty", expressions of this type are also known as variance-based uncertainty relations (VUR). Since Robertson and Schrödinger's papers, a multitude of VURs has been devised; examples include the Fourier-type uncertainty relations of Bohr and Wigner [71], the fractional Fourier-type uncertainty relations of Mustard [233], mixed-states uncertainty relations [75], the angle-angular momentum uncertainty relation of Lévy–Leblond [197] and Carruthers and Nietto [53], the time-energy uncertainty relation of Mandelstam and Tamm [221], Luisell's amplitude-phase uncertainty relation [203], and Synge's three-observable uncertainty relations [301].

2.4.2 *Why do we need Information Theoretic URs?*

Unfortunately, even VURs have many limitations. In fact, the essence of a VUR is to put an upper bound to the degree of concentration of two (or more) probability distributions, or, equivalently impose a lower bound to

the associated uncertainties. While the variance is often a good measure of the concentration of a given distribution, there are many situations where this is not the case. For instance, variance as a measure of concentration is a dubious concept in the case when a distribution contains more than one peak (multimodal distribution). Let me illustrate this with two simple examples [35]:

Example I: Consider two possible situations of a particle in one dimension. *First situation* describes a particle with a *uniform probability density* in a box of total length L, i.e.

$$\varrho \;=\; \begin{cases} 1/L, \text{ inside the box;} \\ 0, \quad \text{outside the box.} \end{cases}$$

Second situation describes a particle localized with *equal* probability densities in two boxes each of length $L/4$, see Fig. 2.9.

$$\varrho \;=\; \begin{cases} 2/L, \text{ inside the box;} \\ 0, \quad \text{outside the box.} \end{cases}$$

Let us now can ask in which situation (F or C) is the uncertainty in

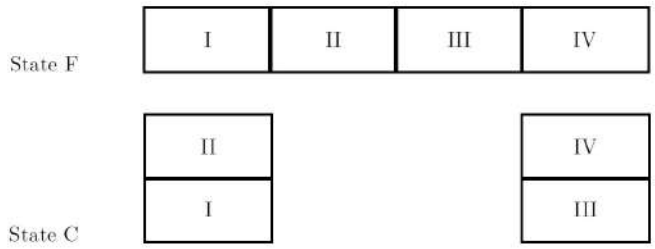


Fig. 2.9: Two states F = flat and C = clustered.

particle's position greater. Intuition would suggest that the uncertainty is greater in the case F since in the case C we know more about particle's position (particle is not in the regions II and III). However, explicit calculation shows that $\Delta x_F = L/\sqrt{12}$ while $\Delta x_C = \sqrt{7/4}L/\sqrt{12}$.

Example II: Consider a particle in one dimension where the probability density is *constant* in two regions I and II separated by a large distance

NL (N is a large number). The region I has the size $L(1-1/N)$ and the distant region II has the size L/N (see Fig. 2.10). *Probability density* is:

$$\varrho \;=\; \begin{cases} 1/L, \text{ in region } \; I; \\ 1/L, \text{ in region } \; II; \\ 0, \quad \text{otherwise.} \end{cases}$$

Note in particular that Δx tends to infinity with increasing N even though

Fig. 2.10: Example II: $\Delta x \sim (L/\sqrt{12})\sqrt{1+12N}$.

the probability of finding the particle in the region I tends to 1.

These two examples clearly illustrate the problem with the standard deviation as a measure of uncertainty. It gets very high contributions from distant regions because these enter with a *large weight*: namely, the distance from the mean value.

Another troublesome feature of VURs appears in the case of finite-dimensional Hilbert spaces, such as the Hilbert space of spin or angular momentum. In such cases the commutator *cannot be* a multiple of a unit operator [115] and hence the uncertainty product, which is bounded from below by the (normed) mean value of the commutator can attain zero minimum even when one of the distributions is not absolutely localized, i.e., even when the value of one of the observables is not precisely known. In such a case the uncertainty is just characterized by the lower bound of the uncertainty product (i.e., by zero) and thus it only says that this product is greater than zero for some states and equal to zero for others. This is, however, true also in classical physics.

There is yet another important limitation of the variance as measure of the concentration of a given distribution. In particular, variance diverges in many distributions even though such distributions are sharply peaked. Notorious examples of this are provided by heavy-tail distributions such as Lévy [84, 196], Weibull [84] or Cauchy–Lorentz distributions [84]. For instance, in the theory of Bright–Wigner shapes it has been known for along

time [206] that the Cauchy–Lorentz distribution can be freely concentrated into an arbitrarily small region by changing its scale parameter, while its standard deviation remains very large or even infinite.

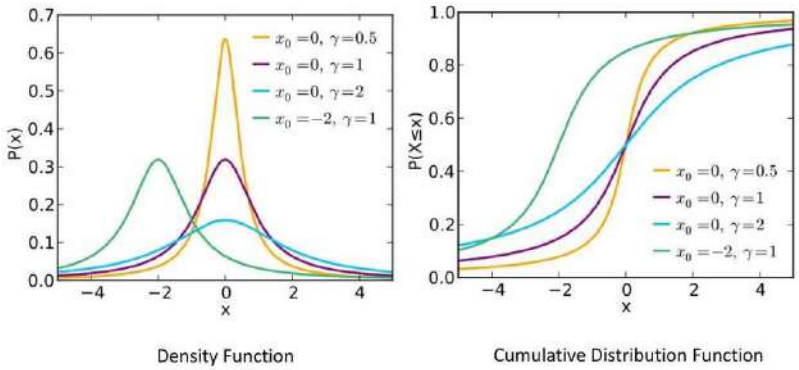


Fig. 2.11: Cauchy PDF can be concentrated into an arbitrarily small region by changing γ, while σ might be even *infinite*.

2.4.3 *Uncertainty relations based on Rényi entropy*

From the previous account we see that there is a need to quantify quantum unpredictability also differently. Among non-variance-based uncertainty relations a particularly prominent role is played by uncertainty relations that are conceptually rooted in information theory. In these cases the uncertainty is quantified in terms of various information measures — entropies, which often provide more stringent bound on concentrations of the probability distributions. The purpose of this section is to give a brief account of some key information-theoretic uncertainty relations (ITUR) and present some new results based on Rényi entropy. In particular, we will see that the system of RE based ITUR's allows to quantify a shape of multimodal distributions and make sense even for heavy tailed distributions.

2.4.3.1 *Discrete probability case*

Mathematical underpinning of most uncertainty relations is in theory of inequalities. For instance, the celebrated Robertson–Schrödinger VUR is based on the Cauchy–Schwarz inequality. Similarly, the Riesz–Thorin in-

equality (also known as the Riesz–Thorin interpolation theorem) [118] allows to prove the following Rényi's ITUR for discrete PDF's [146]:

Theorem 2.3. *Let* $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{C}^n$. *Suppose that* $(\mathcal{L}\boldsymbol{x})_j = \sum_i a_{ij} x_j \equiv (\mathbb{A}\boldsymbol{x})_j$ *and that*

$$\sum_j |(\mathcal{L}\boldsymbol{x})_j|^2 \;\leq\; \sum_j |x_i|^2 \,, \quad \textit{for all } x_i \,.$$

Define $c \equiv \max_{i,j} |a_{ij}|$. *If* $r \in [-1/2, 0]$ *and* $t = -r/(2r+1)$ *and the probability distributions* $\mathcal{P}^{(1)}$ *and* $\mathcal{P}^{(2)}$ *are related by* $\mathcal{L}\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(2)}$ *where* $|x_i| \equiv \xi_i = \sqrt{p_i}$, *then*

$$\mathcal{I}_{1+t}(\mathcal{P}^{(2)}) + \mathcal{I}_{1+r}(\mathcal{P}^{(1)}) \;\geq\; -2\log_2 c \,. \tag{2.64}$$

The proof of this theorem is relegated to Appendix E. Let me now discuss some implications of the inequality (2.64), more detailed discussion can be found in Ref. [146].

We might first observe that since information measures $\mathcal{I}_q(\mathcal{P})$ are always non-negative (for all q), the inequality (2.64) can represent a genuine uncertainty relation only when $c < 1$. On the other hand, for $\mathbb{A} \in SO(n)$ or $SU(n)$ (i.e. for most physically relevant situations) one always has that $c \leq 1$. This is because for such $\mathbb{A}$'s

$$c \;=\; \max_{i,j} |a_{ik}| \;=\; \|\mathbb{A}\|_{\max} \;\leq\; \|\mathbb{A}\|_2 \;=\; \sqrt{\lambda_{\max}(\mathbb{A}^\dagger \mathbb{A})} \;=\; 1 \,. \tag{2.65}$$

The last identity results from the fact[5] that all of eigenvalues of $\mathbb{A} \in SO(n)$ or $SU(n)$ have absolute value 1. Rényi's ITUR (2.64) was originally found by Kraus [184] and Maassen [207].

In the particular case when $r = 0$ (and thus $t = 0$) we get the Shannon entropy based ITUR

$$\mathcal{H}(\mathcal{P}^{(2)}) \;+\; \mathcal{H}(\mathcal{P}^{(1)}) \;\geq\; -2\log_2 c \,. \tag{2.66}$$

A weaker version of this Shannon's ITUR was also earlier proposed by Deutsch [69] and Bialynicky-Birula [36].

Let me now illustrate that the ITUR's (2.64) can provide more stringent bound on concentrations of PDF's than VUR's. To this end let $|\phi\rangle$ be the two-dimensional state of a spin-$\frac{1}{2}$ particle, and let $\hat{S}_x$ and $\hat{S}_z$ be *spin components* in *orthogonal* directions:

$$|S_x\rangle \;\equiv\; \begin{pmatrix} |S_x;+\rangle \\ |S_x;-\rangle \end{pmatrix}, \qquad |S_z\rangle \;\equiv\; \begin{pmatrix} |S_z;+\rangle \\ |S_z;-\rangle \end{pmatrix}. \tag{2.67}$$

[5] One can be even more specific. Since for any matrix $\mathbb{A} \in \mathbb{R}^{n\times n}$ the following inequality holds $\|\mathbb{A}\|_{\max} \;\leq\; \|\mathbb{A}\|_2 \;\leq\; n\|\mathbb{A}\|_{\max}$, which implies that for $\mathbb{A} \in SO(n)$ we have that $c \in [1/n, 1]$.

Because

$$\begin{pmatrix} |S_x;+\rangle \\ |S_x;-\rangle \end{pmatrix} \;=\; \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} |S_z;+\rangle \\ |S_z;-\rangle \end{pmatrix}, \tag{2.68}$$

we have that $c = 1/\sqrt{2}$.

Let now $\mathcal{P} = (p,(1-p)) = (|\langle S_x;+|\phi\rangle|^2, |\langle S_x;-|\phi\rangle|^2)$. Question now stands how this restricts $\mathcal{Q} = (q,(1-q)) = (|\langle S_z;+|\phi\rangle|^2, |\langle S_z;-|\phi\rangle|^2)$

In case of Shannon's ITUR we have

$$\mathcal{H}(\mathcal{P}) \;+\; \mathcal{H}(\mathcal{Q}) \;\geq\; -\log_2 \tfrac{1}{2} \;=\; 1\,, \tag{2.69}$$

which can be equivalently phrased in the form

$$p^p (1-p)^{1-p} \;\leq\; \tfrac{1}{2}\, q^{-q} (1-q)^{q-1}\,. \tag{2.70}$$

A graphical solution of this inequality is depicted at Fig. 2.12.

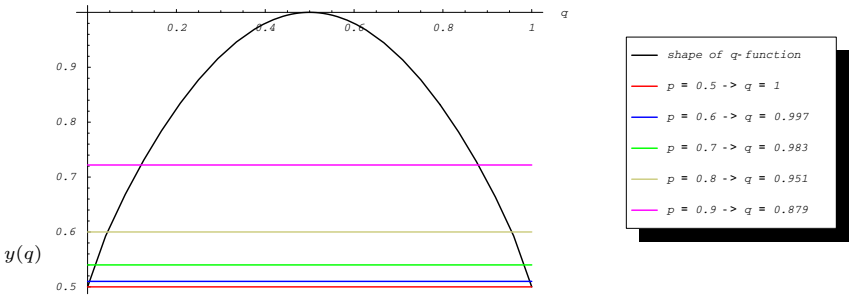


Fig. 2.12: Graphical representation of the inequality (2.70). For fixed value of p the inequality is fulfilled for all q's that lie inside the q-function, i.e., function $y(q) = \frac{1}{2}\, q^{-q}(1-q)^{q-1}$.

Along similar lines I can employ Rényi's ITUR. The most stringent relation between p and q is provided via Rényi's ITUR

$$\mathcal{I}_\infty(\mathcal{P}) \;+\; \mathcal{I}_{1/2}(\mathcal{Q}) \;\geq\; -2\log_2 c \;=\; 1\,, \tag{2.71}$$

which is equivalent to

$$\sqrt{q}\,\sqrt{1-q} \;+\; 1/2 \;\geq\; p\,. \tag{2.72}$$

Again, one can seek for the solution graphically, see Fig. 2.13.

The comparison with the ordinary Schrödinger–Robertson's VUR, can also be made easily. In fact, we have

$$\langle (\triangle S_x)^2 \rangle_\phi \langle (\triangle S_z)^2 \rangle_\phi \;\geq\; \frac{\hbar^2}{4} |\langle S_y \rangle_\phi|^2\,, \tag{2.73}$$

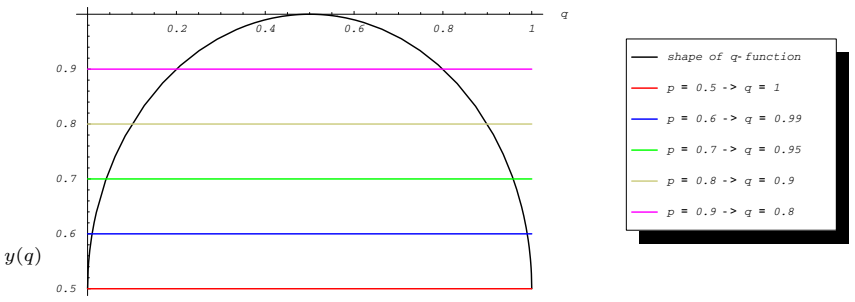


Fig. 2.13: Graphical representation of the inequality (2.72). For fixed value of p the inequality is fulfilled for all q's that lie inside the q-function, i.e., function $y(q) = \sqrt{q}\sqrt{1-q} + 1/2$.

which can be equivalently rewritten as

$$p(1-p) \;\geq\; \tfrac{1}{4}\sin^2(\varphi_+ - \varphi_-)\,, \tag{2.74}$$

where the phase $\varphi_\pm$ is defined as

$$e^{i\varphi_\pm} \;\equiv\; \frac{\langle\phi|S_z;\pm\rangle}{|\langle\phi|S_z;\pm\rangle|}\,. \tag{2.75}$$

Note, that by symmetry the VUR inequality can also be equally written as

$$q(1-q) \;\geq\; \tfrac{1}{4}\sin^2(\tilde{\varphi}_+ - \tilde{\varphi}_-)\,, \tag{2.76}$$

with

$$e^{i\tilde{\varphi}_\pm} \;\equiv\; \frac{\langle\phi|S_x;\pm\rangle}{|\langle\phi|S_x;\pm\rangle|}\,. \tag{2.77}$$

From (2.73) and (2.76) we see that the VUR does not pose any strong restriction between $\mathcal{P}$ and $\mathcal{Q}$. Since the phase factors $\varphi_\pm$ (or $\tilde{\varphi}_\pm$) do not enter the definition of $\mathcal{Q}$ (or $\mathcal{P}$), then for a fixed (but otherwise arbitrary) q the VUR (2.73) can be in principle fulfilled by any $p \in [0.5, 1]$. Of course, if the relative phase is known the restriction between $\mathcal{P}$ and $\mathcal{Q}$ is less trivial. On the other hand, the ITURs discussed above are far more specific in their constrains on values of $\mathcal{P}$ and $\mathcal{Q}$, see Tab. 2.2. From the table we see that for given $\mathcal{P}$, Rényi's ITUR considered improves on Shannon's ITUR. Indeed, for instance, the marginal case $\mathcal{P} = (0.8, 0.2)$ and $\mathcal{Q} = (0.951, 0.049)$ that is allowed by Shannon's ITUR explicitly violates Rényi's ITUR and hence it cannot be realized (ITURs represent necessary conditions). Both Shannon's ITUR and Rényi's ITUR improve on VUR — unless some extra information

p	VUR $q \in$	S-ITUR $q \in$	R-ITUR $q \in$
0.5	$[0.067, 0.933]$	$[0, 1]$	$[0, 1]$
0.6	$[0.067, 0.933]$	$[0.003, 0.997]$	$[0.010, 0.990]$
0.7	$[0.067, 0.933]$	$[0.017, 0.983]$	$[0.042, 0.958]$
0.8	$[0.067, 0.933]$	$[0.049, 0.951]$	$[0.1, 0.9]$
0.9	$[0.067, 0.933]$	$[0.121, 0.879]$	$[0.2, 0.8]$

Table 2.2: Comparison of three uncertainty relations: variance-based uncertainty relation (VUR) with $\tilde{\varphi}_+ - \tilde{\varphi}_- = \pi/6$, Shannon's information uncertainty relation (S-ITUR) and Rényi's information uncertainty relation (R-ITUR) for different values of p. In the respective columns one can see the peakedness of the distribution $\mathcal{Q} = (q, (1-q))$.

about the relative wave-functions phase is provided. In Tab. 2.2 we find that when the relative phase is known, e.g., $\tilde{\varphi}_+ - \tilde{\varphi}_- = \pi/6$, Rényi's ITUR still improves on VUR for values $p = 0.9$ and $p = 0.8$ while Shannon's ITUR improves over VUR only for $p = 0.9$.

In its essence the above ITURs represent non-linear constraining relationships between two distributions. Rather than trying to find the best set of parameters in URs that can to put a better upper bound to the degree of concentration of the distributions involved, I prefer to view Rényian ITURs as an infinite (one parameter) class of constraining relations that can be equivalently understood as constraining relations among higher-order moments. So, by their very formulations ITURs go beyond VUR. I will illustrate this point in more detail in the following section

2.4.3.2 *Continuous probability case*

Particularly important class of ITURs for continuous probabilities (or PDFs) are Fourier-transform based entropic uncertainty relations. The key inequality that is used in this context is based on the following Beckner–Babebko's theorem [10, 27]:

Theorem 2.4 (Beckner–Babebko's theorem). *Let*

$$f^{(2)}(\boldsymbol{x}) \equiv \hat{f}^{(1)}(\boldsymbol{x}) = \int_{\mathbb{R}^D} e^{2\pi i \boldsymbol{x}\cdot\boldsymbol{y}}\ f^{(1)}(\boldsymbol{y})\ d\boldsymbol{y}\,,$$

then for $p \in [1,2]$ *we have*

$$\|\hat{f}\|_{p'} \;\leq\; \frac{|p^{D/2}|^{1/p}}{|(p')^{D/2}|^{1/p'}}\;\|f\|_p\,, \tag{2.78}$$

or, equivalently

$$|(p')^{D/2}|^{1/p'}\|f^{(2)}\|_{p'} \;\leq\; |p^{D/2}|^{1/p}\|f^{(1)}\|_p\,.$$

Here, p *and* p' *are the usual Hölder conjugates (i.e.,* $p' \in [2,\infty)$*). For any* $X \in \ell^p(\mathbb{R}^D)$ *the* p*-norm* $\|X\|_p$ *is defined as*

$$\|X\|_p = \left(\int_{\mathbb{R}^D} |X(\boldsymbol{y})|^p \; d\boldsymbol{y}\right)^{1/p}.$$

Due to symmetricity of the Fourier transform also the reverse inequality holds:

$$\|f\|_{p'} \;\leq\; \frac{|p^{D/2}|^{1/p}}{|(p')^{D/2}|^{1/p'}}\;\|\hat{f}\|_p\,. \tag{2.79}$$

In connection with the Beckner–Babenko theorem it is important to mention that Lieb proved in Ref. [199] that the inequalities (2.78)-(2.79) are saturated (for all p) only for Gaussian functions. I postpone the proof of the Beckner–Babebko's theorem to Appendix F.

To proceed, I first introduce the concept of Rényi's entropy power (REP). This is defined as the solution of the equation [146, 158]

$$\mathcal{I}_q\left(\mathcal{X}\right) \;=\; \mathcal{I}_q\left(\sqrt{N_q^R(\mathcal{X})}\cdot\mathcal{Z}_G\right), \tag{2.80}$$

where $\{\mathcal{Z}_G\}$ represents a *Gaussian random vector* with zero mean and unit covariance matrix. So, $N_p^R(\mathcal{X})$ denotes the variance of a would be Gaussian distribution that has the same Rényi information content as the random vector $\{\mathcal{X}_i\}$. Expression (2.80) was studied in [95, 146, 158] where it was shown that the only class of solutions of (2.80) is

$$N_q^R(\mathcal{X}) \;=\; \frac{1}{2\pi}p^{-q'/q}\exp\left(\frac{2}{D}\,\mathcal{I}_q(\mathcal{X})\right), \tag{2.81}$$

with $1/p + 1/p' = 1$ and $p \in \mathbb{R}^+$. In addition, when $p \to 1_+$ one has $N_p^R(\mathcal{X}) \to N(\mathcal{X})$, where $N(\mathcal{X})$ is Shannon's entropy power introduced by Shannon himself in Ref. [286]. In this latter case, one can use the *asymptotic equipartition property* [62, 153] to identify $N(\mathcal{X})$ with "typical size" of a state set, which in the present context is the effective support set size for a random vector. In passing, we may observe that from (2.81) it follows that $N_q^R(\sigma\mathcal{Z}_G) = \sigma^2$, i.e. for Gaussian processes the REP is simply the variance

σ^2. In the case where $\mathcal{Z}_G^K$ represents a random Gaussian vector of zero mean and covariance matrix K_{ij}, then $N_q^R(\mathcal{Z}_G^K) = [\det(K_{ij})]^{1/D} \equiv |K|^{1/D}$.

It should be also noted that the form of the REP expressed in (2.81) is not universally accepted version. In a number of works, it is defined merely as an exponent of RE, see, e.g., [67, 257]. My motivation for the form (2.81) is twofold: first, it has a clear interpretation in terms of variances of Gaussian distributions and, second, it leads to simpler formulas, cf. e.g., Eq. (2.88).

The passage to quantum mechanical UR is now quite straightforward. First, we realize that, in quantum mechanics, the Fourier conjugate wave functions are related via two reciprocal relations

$$\begin{aligned} \psi(\boldsymbol{x}) &= \int_{\mathbb{R}^D} e^{i\boldsymbol{p}\cdot\boldsymbol{x}/\hbar}\, \hat{\psi}(\boldsymbol{p})\, \frac{d\boldsymbol{p}}{(2\pi\hbar)^{D/2}}\,, \\ \hat{\psi}(\boldsymbol{p}) &= \int_{\mathbb{R}^D} e^{-i\boldsymbol{p}\cdot\boldsymbol{x}/\hbar}\, \psi(\boldsymbol{x})\, \frac{d\boldsymbol{x}}{(2\pi\hbar)^{D/2}}\,. \end{aligned} \tag{2.82}$$

Plancherel (or Riesz–Fischer) equality then implies that $\|\psi\|_2 = \|\hat{\psi}\|_2 = 1$. Let me define new functions, namely

$$\begin{aligned} f^{(2)}(\boldsymbol{x}) &= (2\pi\hbar)^{D/4}\psi(\sqrt{2\pi\hbar}\,\boldsymbol{x})\,, \\ f^{(1)}(\boldsymbol{p}) &= (2\pi\hbar)^{D/4}\hat{\psi}(\sqrt{2\pi\hbar}\,\boldsymbol{p})\,. \end{aligned} \tag{2.83}$$

The factor $(2\pi\hbar)^{D/4}$ ensures that also the new functions are normalized (in sense of $\|\ldots\|_2$) to unity. With these we will have the same structure of the Fourier transform as in the Beckner–Babenko inequality. Beckner–Babenko inequality can be then rewritten as

$$\left[\left(\frac{q'}{2\pi\hbar}\right)^D\right]^{1/q'} \||\psi|^2\|_{q'/2} \;\le\; \left[\left(\frac{q}{2\pi\hbar}\right)^D\right]^{1/q} \||\hat{\psi}|^2\|_{q/2}\,. \tag{2.84}$$

This is equivalent to

$$\begin{aligned} &\left[\left(\frac{q'}{2\pi\hbar}\right)^D\right]^{1/q'} \exp\left[\frac{2(1-q'/2)}{q'}\,\mathcal{I}_{q'/2}\left(|\psi|^2\right)\right] \\ &\qquad\le\; \left[\left(\frac{q}{2\pi\hbar}\right)^D\right]^{1/q} \exp\left[\frac{2(1-q/2)}{q}\,\mathcal{I}_{q/2}\left(|\hat{\psi}|^2\right)\right]. \end{aligned} \tag{2.85}$$

Now I take power $q/(D(1-q/2))$ of both left and right side and use the

fact that $2/q - 1 = 1 - 2/q'$. This gives

$$\left(\frac{q}{2\pi\hbar}\right)^{1/(1-q/2)} \exp\left[\frac{2}{D}\,\mathcal{I}_{q/2}\left(|\hat{\psi}|^2\right)\right] \times \left(\frac{q'}{2\pi\hbar}\right)^{1/(1-q'/2)} \exp\left[\frac{2}{D}\,\mathcal{I}_{q'/2}\left(|\psi|^2\right)\right] \;\geq\; 1\,. \quad (2.86)$$

This is equal to (use that for Hölder double one has $1/(1-q/2) + 1/(1-q'/2) = 2$)

$$\underbrace{\frac{1}{2\pi}\left(\frac{q}{2}\right)^{1/(1-q/2)} \exp\left[\frac{2}{D}\,\mathcal{I}_{q/2}\left(|\hat{\psi}|^2\right)\right]}_{N^R_{q/2}(|\hat{\psi}|^2)} \times \underbrace{\frac{1}{2\pi}\left(\frac{q'}{2}\right)^{1/(1-q'/2)} \exp\left[\frac{2}{D}\,\mathcal{I}_{q'/2}\left(|\psi|^2\right)\right]}_{N^R_{q'/2}(|\psi|^2)} \;\geq\; \frac{\hbar^2}{4}\,. \quad (2.87)$$

Hence REP-based Beckner–Babenko inequality acquires in the QM setting a simple form

$$N^R_{q/2}(|\hat{\psi}|^2) N^R_{q'/2}(|\psi|^2) \;\geq\; \frac{\hbar^2}{4}\,. \quad (2.88)$$

This represents an infinite tower of mutually distinct (generally irreducible) REP-URs [158].

In connection with (2.88) one might observe that the conventional Robertson–Schrödinger URs [271, 281]) and Shannon differential entropy based URs (e.g., Hirschman or Białynicki-Birula URs [36, 128]) naturally appear as special cases in this hierarchy. Indeed, when $\mathcal{X}$ is a random Gaussian vector, then $\mathcal{Y}$ is also Gaussian and (2.88) reduces to

$$|K_{\mathcal{X}}|^{1/D} |K_{\mathcal{Y}}|^{1/D} \;=\; \frac{\hbar^2}{4}\,. \quad (2.89)$$

The equality follows from the saturation of the Beckner–Babenko inequality (2.84) by Gaussian functions.

By assuming that a PDF has a finite covariance matrix $(K_{\mathcal{X}})_{ij}$ then important inequalities hold, namely

$$N(\mathcal{X}) \;\leq\; |K_{\mathcal{X}}|^{1/D} \;\leq\; \sigma^2_{\mathcal{X}}\,, \quad (2.90)$$

with equality in the first inequality if and only if $\mathcal{X}$ is a Gaussian vector, and in the second if and only if $\mathcal{X}$ has covariance matrix that is proportional to

the identity matrix. The proof of (2.90) is based on the non-negativity of the Kullback–Leibler divergence and can be found in Appendix D. Inequality (2.88) together with (2.90) immediately gives

$$\sigma_{\mathcal{X}}^2\sigma_{\mathcal{Y}}^2 \geq |K_{\mathcal{X}}|^{1/D}|K_{\mathcal{Y}}|^{1/D} \;\geq\; N(\mathcal{X})N(\mathcal{Y}) \;\geq\; \frac{\hbar^2}{4}\,, \tag{2.91}$$

which saturates only for Gaussian (respective white) random vectors $\mathcal{X}$ and $\mathcal{Y}$. Note, that when $(K_{\mathcal{X}})_{ij}$ and $(K_{\mathcal{Y}})_{ij}$ exist then (2.91) automatically implies the conventional Robertson–Schrödinger variance-based UR.

Second, the information–theoretic entropies enter quantum information theory typically in three distinct ways: *a*) as a measure of the quantum information content (e.g., how many qubits are needed to encode the message without loss of information), *b*) as a measure of the classical information content (e.g., amount of information in bits that can be recovered from the quantum system) and *c*) to quantify the entanglement of pure and mixed bipartite quantum states. Logarithms in base 2 are then typically used because in quantum information, because one quantifies entropy in bits and qubits (rather than nats). This in turn also modifies Rényi's EP as

$$\frac{1}{2\pi}p^{-p'/p}e^{\left(\frac{2}{D}\cdots\right)} \;\mapsto\; \frac{1}{2\pi}p^{-p'/p}\,2^{\left(\frac{2}{D}\cdots\right)}\,. \tag{2.92}$$

Third, though the most prominent example of the Fourier transform (2.82) is that between configuration and momentum space wave functions, one can use for the purpose of the ITUR (2.92) also other Fourier transform duals, such as the angular momentum and angle or conjugate quandratures [158].

2.4.3.3 *Information distribution*

Since the variance-based UR is implied by the Shannon EPUR alone, a natural question arises; in what sense is the general set of inequalities (2.88) more informative than the special case $q = q' = 2$? To aid our intuition and, furthermore, to show the conceptual underpinning for REP-URs (2.88) I first start with the concept of *information distributions* (or better *information PDF*).

Let $\mathcal{F}(\boldsymbol{x})$ be the PDF for the random variable $\mathcal{X}$. I define the *information random variable* $i_{\mathcal{X}}(\mathcal{X})$ so that $i_{\mathcal{X}}(\boldsymbol{x}) = \log_2 1/\mathcal{F}(\boldsymbol{x})$. In other word $i_{\mathcal{X}}(\boldsymbol{x})$ represents the information in $\boldsymbol{x}$ with respect to $\mathcal{F}(\boldsymbol{x})$. In this connection it is expedient to introduce the cumulative distribution function for $i_{\mathcal{X}}(\mathcal{X})$ as

$$\wp(y) \;=\; \int_{-\infty}^{y} d\wp(i_{\mathcal{X}}) \;=\; \int_{\mathbb{R}^D} \mathcal{F}(\boldsymbol{x})\theta(\log_2 \mathcal{F}(\boldsymbol{x}) + y)d\boldsymbol{x}\,. \tag{2.93}$$

Such $\wp(y)$ thus represents the probability that the random variable $i_{\mathcal{X}}(\mathcal{X})$ is equal or less than y. I have denoted the corresponding probability measure as $d\wp(i_{\mathcal{X}})$. Taking the Laplace transform of both sides of (2.93), we get

$$\mathcal{L}\{\wp\}(s) \;=\; \int_{\mathbb{R}^D} \mathcal{F}(\boldsymbol{x})\, \frac{e^{s\log_2 \mathcal{F}(\boldsymbol{x})}}{s}\, d\boldsymbol{x} \;=\; \frac{\mathbb{E}\left[e^{s\log_2 \mathcal{F}}\right]}{s}\,, \tag{2.94}$$

where $\mathbb{E}\left[\cdots\right]$ denotes the mean value with respect to $\mathcal{F}$. By assuming that $\wp(x)$ is smooth then the PDF associated with $i_{\mathcal{X}}(\mathcal{X})$ — so-called *information PDF*, is

$$g(y) \;=\; \frac{d\wp(y)}{dy} \;=\; \mathcal{L}^{-1}\left\{\mathbb{E}\left[e^{s\log_2 \mathcal{F}}\right]\right\}(y)\,. \tag{2.95}$$

By setting $s = (p-1)\log 2$ I get

$$\mathcal{L}\{g\}(s=(p-1)\log 2) \;=\; \mathbb{E}\left[2^{(1-p)i_{\mathcal{X}}}\right]. \tag{2.96}$$

The mean is taken here with respect to the PDF g. Eq.(2.96) can be written also explicitly as

$$\int_{\mathbb{R}^D} d\boldsymbol{x}\, \mathcal{F}^p(\boldsymbol{x}) \;=\; \int_{\mathbb{R}} g(y) 2^{(1-p)y} dy\,. \tag{2.97}$$

Note that when $\mathcal{F}^p$ is integrable for $p \in [1,2]$ then (2.97) ensures that the moment-generating function for $g(x)$ PDF exists. So, in particular, the moment-generating function exists when $\mathcal{F}(\boldsymbol{x})$ represents Lévy α-stable distributions, including the heavy-tailed stable distributions (i.e, PDFs with the Lévy stability parameter $\alpha \in (0,2]$). The same holds for $\hat{\mathcal{F}}$ and $p' \in [2,\infty)$ due to the Beckner–Babenko theorem [10, 27, 146] (cf. also Appendix F).

2.4.3.4 *Reconstruction theorem*

Since $\mathcal{L}\{g\}(s)$ from (2.96) is the *moment-generating function* of the random variable $i_{\mathcal{X}}(\mathcal{X})$ one can find all moments of the PDF $g(x)$ (if they exist) by taking the derivatives of $\mathcal{L}\{g\}$ with respect to s. From conceptual reasons it is often more convenient to work with cumulants rather than moments. Using the fact that the *cumulant-generating function* is simply the logarithm of the moment-generating function, we see from (2.97) that the differential RE is a reparametrized version of the cumulant generating function of the information random variable $i_{\mathcal{X}}(\mathcal{X})$. In fact, from (2.96) I can write

$$\mathcal{I}_p(\mathcal{X}) \;=\; \frac{1}{(1-p)} \log_2 \mathbb{E}\left[2^{(1-p)i_{\mathcal{X}}}\right]. \tag{2.98}$$

To understand the meaning of REP-URs I begin with the cumulant expansion (2.98), i.e.

$$p\mathcal{I}_{1-p}(\mathcal{X}) \;=\; \log_2 e \sum_{n=1}^{\infty} \frac{\kappa_n(\mathcal{X})}{n!} \left(\frac{p}{\log_2 e}\right)^n, \tag{2.99}$$

where $\kappa_n(\mathcal{X}) \equiv \kappa_n(i_{\mathcal{X}})$ denotes the n-th cumulant of the information random variable $i_{\mathcal{X}}(\mathcal{X})$ (in units of $bits^n$). Let us note that

$$\begin{aligned} \kappa_1(\mathcal{X}) \;&=\; \mathbb{E}\left[i_{\mathcal{X}}(\mathcal{X})\right] \;=\; \mathcal{H}(\mathcal{X}), \\ \kappa_2(\mathcal{X}) \;&=\; \mathbb{E}\left[i_{\mathcal{X}}(\mathcal{X})^2\right] \;-\; \left(\mathbb{E}\left[i_{\mathcal{X}}(\mathcal{X})\right]\right)^2, \end{aligned} \tag{2.100}$$

i.e., they represent the entropy and *varentropy*, respectively. By employing the identity

$$\mathcal{I}_{1-p}(\mathcal{X}) = \frac{D}{2} \log_2 \Big[2\pi(1-p)^{-1/p} N_{1-p}(\mathcal{X})\Big], \tag{2.101}$$

I can rewrite (2.99) in the form

$$\begin{aligned} &\log_2 \left[N_{1-p}(\mathcal{X})\right] \\ &\quad= \log_2 \left[\frac{(1-p)^{1/p}}{2\pi}\right] + \frac{2}{D} \sum_{n=1}^{\infty} \frac{\kappa_n(\mathcal{X})}{n!} \left(\frac{p}{\log_2 e}\right)^{n-1}. \end{aligned} \tag{2.102}$$

I turn, from (2.102) one can see that

$$\begin{aligned} \kappa_n(\mathcal{X}) \;&=\; \frac{nD}{2} (\log_2 e)^{n-1} \frac{d^{n-1} \log_2 \left[N_{1-p}(\mathcal{X})\right]}{dp^{n-1}}\bigg|_{p=0} \\ &\quad+\; \frac{D}{2} (\log_2 e)^n \left[(n-1)! + \delta_{1n} \log 2\pi\right], \end{aligned} \tag{2.103}$$

(here δ_{ij} is the Kronecker delta) which, in terms of the Grünwald–Letnikov derivative formula [275], allows us to write

$$\begin{aligned} \kappa_n(\mathcal{X}) \;&=\; \lim_{\Delta \to 0} \frac{nD}{2} \frac{(\log_2 e)^n}{\Delta^{n-1}} \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \log \left[N_{1+k\Delta}(\mathcal{X})\right] \\ &\quad+\; \frac{D}{2} (\log_2 e)^n \left[(n-1)! \;+\; \delta_{1n} \log 2\pi\right]. \end{aligned} \tag{2.104}$$

So, in order to determine the first m cumulants of $i_{\mathcal{X}}(\mathcal{X})$ we need to know all $N_1, N_{1+\Delta}, \ldots, N_{1+(m-1)\Delta}$ entropy powers. In practice Δ corresponds to a characteristic resolution scale for the entropy index which is typically of the order 10^{-2}.

When all cumulants exist then the problem of recovering the underlying PDF for $i_{\mathcal{X}}(\mathcal{X})$ is equivalent to the *Stieltjes* moment problem [260]. In this

connection, there are a number of ways to proceed; the PDF in question can be reconstructed e.g., in terms of sums involving orthogonal polynomials (e.g., the Gram–Charlier A series or the Edgeworth series [327]), the inverse Mellin transform [336] or via various maximum entropy techniques [303]. Pertaining to this, the theorem of Marcinkiewicz [204] implies that there are no PDFs for which $\kappa_m = \kappa_{m+1} = \ldots = 0$ for some $m \geq 3$. In other words, the cumulant generating function cannot be a finite-order polynomial of degree greater than 2. The important exceptions, and indeed the only exceptions to Marcinkiewicz's theorem, are the *Gaussian* PDFs which can have the first two cumulants nontrivial and $\kappa_3 = \kappa_4 = \ldots = 0$. Thus, apart from the special case of Gaussian PDFs where only N_1 and $N_{1+\Delta}$ are needed, one needs to work with as many entropy powers $N_{1+k\Delta}, k \in \mathbb{N}$ as possible to receive as much information as possible about the structure of the underlying PDF. In theory, the whole infinite tower of REP-URs would be required to uniquely specify a system's information PDF. From (2.102) and (2.104) we see that knowledge of N_1 corresponds to $\kappa_1(\mathcal{X}) = \mathcal{H}(\mathcal{X})$ while $N_{1+\Delta}$ further determines κ_2, i.e. the varentropy. Since N_1 is involved [through (2.104)] in the determination of all cumulants, it is the most important entropy power in the tower.

I should stress, that the focus of the presented reconstruction theorem is on cumulants κ_n which can be directly used for a shape estimation of $g(x)$ but not $\mathcal{F}(\boldsymbol{x})$. However, by knowing $g(y)$ we have a complete "information scan" of $\mathcal{F}(\boldsymbol{x})$. Such an information scan is, however, not unique, indeed two PDFs that are rearrangements of each other — *equimeasurable* PDFs have identical both $\wp(y)$ and $g(y)$. Even though equimeasurable PDFs cannot be distinguished via their entropy powers, they can be, as a rule, distinguished via their respective momentum-space PDFs and associated entropy powers. So, the information scan has a tomographic flavor to it. It should be noted that from the multi-peak structure of $g(y)$ one can determine the *number* and *height* of the stationary points (for moro details see [158]). These are invariant characteristics of a given family of equimeasurable PDFs.

2.4.3.5 *Information scan of the state PDF*

Let me now briefly illustrate how one can in practice reconstruct the information distribution $g(x)$ from REPs. I will do this by using the (generalized) Gram–Charlier A expansion [327]. At the same time I should stress that other — often more efficient methods — are also available [327].

Let κ_n be cumulants obtained from entropy powers and let $G(x)$ be some reference PDF whose cumulants are γ_k. The information PDF $g(x)$ can be then written as [327]

$$g(x) \;=\; \exp\left[\sum_{k=1}^{\infty}(\kappa_k-\gamma_k)(-1)^k\frac{(d^k/dx^k)}{k!}\right]G(x)\,. \tag{2.105}$$

With the hindsight I choose the reference PDF $G(x)$ to be a shifted gamma PDF, i.e.

$$G(x) \;\equiv\; \mathcal{G}(x|a,\alpha,\beta) \;=\; \frac{e^{-(x-a)/\beta}(x-a)^{\alpha-1}}{\beta^{\alpha}\Gamma[\alpha]}\,, \tag{2.106}$$

with $a<x<\infty,\ \beta>0,\ \alpha>0$. In doing so, I have implicitly assumed that the $\mathcal{F}(y)$ PDF is in the first approximation equimeasurable with the Gaussian PDF [cf. Eqs. (2.93) and (2.95)]. To reach a corresponding matching we should choose $a=\log_2(2\pi\sigma^2)/2$, $\alpha=1/2$ and $\beta=\log_2 e$. Using the fact that [246]

$$(\beta)^{k+1/2}\,\frac{d^k\,\mathcal{G}(x|a,1/2,\beta)}{k!\,dx^k} \;=\; \left(\frac{x-a}{\beta}\right)^{-k} L_k^{(-1/2-k)}\left(\frac{x-a}{\beta}\right)\mathcal{G}(x|a,1/2,\beta)\,, \tag{2.107}$$

(where L_k^{δ} is an associated Laguerre polynomial of order k with parameter δ) and given that $\kappa_1=\gamma_1=\alpha\beta+a=\log_2(2\pi\sigma^2 e)/2$, and $\gamma_k=\alpha\beta^k=(\log_2 e)^k/2$ for $k>1$ I can write (2.105) as

$$g(x) \;=\; \mathcal{G}(x|a,1/2,\beta)\left[1 \;+\; \frac{(\kappa_2-\gamma_2)}{\beta^{1/2}\,(x-a)^2}\;L_2^{(-3/2)}\left(\frac{x-a}{\beta}\right)\right.$$

$$\left. -\;\frac{(\kappa_3-\gamma_3)}{\beta^{1/2}\,(x-a)^3}\;L_3^{(-7/2)}\left(\frac{x-a}{\beta}\right) \;+\; \cdots\right]. \tag{2.108}$$

If needed, one can use a relationship between the moments and the cumulants (Faà di Bruno's formula [204]) to recast the expansion (2.108) into more familiar language. For the Gram–Charlier A expansion various formal convergence criteria exist (see, e.g., [327]). In particular, the expansion for nearly Gaussian equimeasurable PDFs $\mathcal{F}(y)$ converges quite rapidly and the series can be truncated fairly quickly. Since in this case one needs fewer κ_k's in order to determine the information PDF $g(x)$, only EPs in the small neighborhood of the index 1 will be needed. On the other hand, the further the $\mathcal{F}(y)$ is from Gaussian (e.g., heavy-tailed PDFs) the higher orders of κ_k will be required to determine $g(x)$, and hence a wider neighborhood of the index 1 will be needed for EPs. Some explicit information scans based on formula (2.108) were presented in Refs. [147, 158].

2.4.3.6 *Note on Tsallis entropy powers*

Let me now briefly comment on the entropy powers associated with *Tsallis differential entropy*, i.e. the differential entropy of the form [314]

$$\mathcal{S}_q(\mathcal{F}) \;=\; \frac{1}{(1-q)}\left[\int_{\mathbb{R}^D}\left(\mathcal{F}^q(\boldsymbol{x})-\mathcal{F}(\boldsymbol{x})\right)d\boldsymbol{x}\right], \tag{2.109}$$

where, the PDF $\mathcal{F}(\boldsymbol{x})$ is associated with a D-dimensional random vector $\{\mathcal{X}_i\}$.

By analogy with the RE case, I define the Tsallis entropy power $N_q^T(\mathcal{X})$ as the solution of the equation [cf. Eq. (2.80)]

$$\mathcal{S}_q\left(\mathcal{X}\right) \;=\; \mathcal{S}_q^T\left(\sqrt{N_q^T(\mathcal{X})}\cdot\mathcal{Z}^G\right). \tag{2.110}$$

N_q^T can be easily derived from the scaling property

$$\mathcal{S}_q(a\mathcal{X}) \;=\; \mathcal{S}_q(\mathcal{X}) \;\oplus_q\; \ln_q|a|^D, \tag{2.111}$$

where $a\in\mathbb{R}$ and the q-deformed sum and q-logarithm are defined in (1.62) and (1.58), respectively. Relation (2.111) results from the fact that

$$\begin{aligned}\mathcal{S}_q(a\mathcal{X}) \;&=\; \frac{1}{1-q}\left[\int_{\mathbb{R}^D}d^D\boldsymbol{y}\left(\int_{\mathbb{R}^D}d^D\mathbf{x}\,\delta(\mathbf{y}-a\mathbf{x})\mathcal{F}(\boldsymbol{x})\right)^q-1\right]\\ &=\; \frac{1}{1-q}\left[|a|^{D(1-q)}\int_{\mathbb{R}^D}d^D\boldsymbol{y}\,\mathcal{F}^q(\boldsymbol{y})-1\right]\\ &=\; |a|^{D(1-q)}\,\mathcal{S}_q(\mathcal{X}) \;+\; \ln_q|a|^D \;=\; \mathcal{S}_q(\mathcal{X}) \;\oplus_q\; \ln_q|a|^D.\end{aligned} \tag{2.112}$$

I will further use the simple result

$$\mathcal{S}_q(\mathcal{Z}_G) \;=\; \ln_q(2\pi q^{q'/q})^{D/2}. \tag{2.113}$$

Here q and q' is a Hölder double, i.e. $1/q+1/q'=1$ with $q,q'\in\mathbb{R}^+$. If I now combine (2.110)-(2.113) I obtain

$$\begin{aligned}\mathcal{S}_q(\mathcal{X}) \;&=\; \ln_q(2\pi q^{q'/q})^{D/2} \;\oplus_q\; \ln_q(N_q^T)^D/2\\ &=\; \ln_q(2\pi q^{q'/q}N_q^T)^{D/2}.\end{aligned} \tag{2.114}$$

This directly gives

$$\begin{aligned}N_q^T(\mathcal{X}) \;&=\; \frac{1}{2\pi}q^{-q'/q}\left[\exp_q\left(\mathcal{S}_q(\mathcal{X})\right)\right]^{2/D}\\ &=\; \frac{1}{2\pi}q^{-q'/q}\exp_{1-(1-q)D/2}\left(\frac{2}{D}\,\mathcal{S}_q(\mathcal{X})\right).\end{aligned} \tag{2.115}$$

In addition, when $q \to 1_+$ one has

$$\lim_{q\to 1} N_q^T(\mathcal{X}) \;=\; \frac{1}{2\pi e}\exp\left(\frac{2}{D}H(\mathcal{X})\right) \;=\; N(\mathcal{X})\,, \tag{2.116}$$

where $N(\mathcal{X})$ is the conventional Shannon entropy power (D.11).

In connection with Tsallis EP one might notice that Rényi's EP (considering RE in nats) can be re-written as

$$\begin{aligned} N_q(\mathcal{X}) \;&=\; \frac{1}{2\pi}q^{-q'/q}\exp\left(\frac{2}{D}\,\mathcal{I}_q(\mathcal{X})\right) \\ &=\; \frac{1}{2\pi}q^{-q'/q}\left(\int d^D\mathbf{x}\,\mathcal{F}^q(\boldsymbol{x})\right)^{2/(D(1-q))} \\ &=\; \frac{1}{2\pi}q^{-q'/q}\left[e_q^{\mathcal{S}_q^T(\mathcal{X})}\right]^{2/D} \;=\; N_q^T(\mathcal{X})\,. \end{aligned} \tag{2.117}$$

Here I have used

$$\begin{aligned} \left(\int d^D\mathbf{x}\,\mathcal{F}^q(\boldsymbol{x})\right)^{1/(1-q)} \;&=\; \left[(1-q)\mathcal{S}_q^T(\mathcal{X}) \;+\; 1\right]^{1/(1-q)} \\ &=\; e_q^{\mathcal{S}_q^T(\mathcal{X})}\,. \end{aligned} \tag{2.118}$$

So, we have obtained that Rényi and Tsallis EPs coincide with each other. Consequently, Rényi's EPI (2.88) can be equivalently written in the form

$$N_{p/2}^T(\mathcal{X})N_{q/2}^T(\mathcal{Y}) \;\geq\; \frac{\hbar^2}{4}\,. \tag{2.119}$$

Though URs (2.119) are quite interesting from a mathematical point of view, it is, however, not clear how they could be practically utilized in the estimation theory as there is no obvious operational meaning associated with Tsallis entropy (e.g., there is no coding theorem for Tsallis entropy). On the other hand, Tsallis entropy is important concept in the description of entanglement [328]. For instance, Tsallis entropy of order 2 (also known as linear entropy) directly quantifies state purity [28].

Appendix C

Reprinted papers on Rényi entropy and its applications

[144] P. Jizba and T. Arimitsu, *The world according to Réenyi: thermodynamics of multifractal systems*, Annals of Physics **321** (2004) 17.

[151] P. Jizba, H. Kleinert and M. Shefaat, *Rényi's information transfer between financial time series*, Physica A **391** (2012) 2971.

[158] P. Jizba, Y. Ma, A. Hayes and J.A. Dunningham *One-parameter class of uncertainty relations based on entropy power*, Phys. Rev. E **93** (2016) 060104(R).

[147] P. Jizba and J. Dunningham and M. Prokš *From Rényi Entropy Power to Information Scan of Quantum States*, Entropy **23** (2021) 334.

Appendix D

Some useful relations

D.1 Gibbs inequality and mutual information

Gibbs' inequality is a statement in information theory about Shannon's entropy of a discrete probability distribution. The essence of the inequality is that the *mutual information* between X and Y is non-negative. Let me recall first that mutual information between two random variables X and Y is defined as:

$$\begin{aligned} I(X;Y) = H(X) \; - \; H(X|Y) \; &= \; H(Y) \; - \; H(Y|X) \\ &= \sum_{x \in X,\, y \in Y} p(x,y) \log_2 \frac{p(x,y)}{p(x)q(y)} \,. \end{aligned} \tag{D.1}$$

Gibbs' inequality states that $I(X;Y) \geq 0$ and proof of this is based on Jensen's inequality for concave functions. In particular, one has that for a random variable X $\langle \log_2 X \rangle \leq \log_2 \langle X \rangle$ with equality iff all states of X are identical, i.e., $x_1 = x_2 = \ldots = x_m$.

Proof of Gibbs inequality:

$$\begin{aligned} I(X;Y) \; &= \; - \sum_{x \in X,\, y \in Y} p(x,y) \log_2 \frac{p(x)q(y)}{p(x,y)} \\ &\geq \; - \log_2 \left(\sum_{x \in X,\, y \in Y} p(x)q(y) \right) \; = \; \log_2 1 \; = \; 0 \,. \end{aligned} \tag{D.2}$$

This implies that $H(X) \geq H(X|Y)$ with equality iff $p(x,y) = p(x)q(y)$ for all $x \in X$ and $y \in Y$, i.e. when the two random variables are independent. □

In thermodynamical framework Gibbs' inequality represents a particular variant of the 2nd law of thermodynamics. This can be understood in the sense that in the course of its evolution a large system loses memory of its boundary conditions (represented here by the random variable Y), e.g., due to chaotic behavior of its parts, and thus inevitably increases its entropy.

D.2 Gibbs inequality and relative entropy

Mutual information is a specific realization of *relative entropy* (or the Kullback–Leibler divergence in Shannon's framework) from the product of the marginal distributions, $p_X \cdot p_Y$, to the joint distribution $p_{(X,Y)}$. In particular one has the relation

$$I(X;Y) \;\equiv\; I_1(X;Y) \;=\; D_{KL}(p_{(X,Y)}||p_X \cdot p_Y)\,. \tag{D.3}$$

Gibbs inequality is often more generally formulated in terms of relative entropies in which case it simply states that relative entropy is always non-negative. To see how this comes about let us take X to be random variable with two possible distributions $p_X = p(x)$ (known as reference or prior distribution) and $q_X = q(x)$ (known as updated or posterior distribution). The relative entropy, or equivalently *gain of information* when replacing p_X by q_X, is in Shannon's case

$$\begin{aligned} I_1(q_X||p_X) \;=\; D_{KL}(q_X||p_X) \;&=\; \sum_{x\,\in\,X} q(x)\log_2\frac{q(x)}{p(x)} \\ &\geq\; -\log_2 \sum_{x\,\in\,X} p(x) \;=\; 0\,, \end{aligned} \tag{D.4}$$

where the inequality follows from Jensen's inequality for the concave function $\log_2 x$. Clearly, the relative entropy is zero iff $p_X = q_X$. Since relative entropy represents the information gained by updating from p_X to q_X, zero value of relative entropy means that no information is gained when the updating distribution is the same as the reference one.

It is quite interesting to note that the *relative Rényi entropy* (i.e., relative entropy associated with information measure of the order q ($q \neq 1$)) is

also non-negative. Indeed, I can write

$$\begin{aligned} I_q(q_X||p_X) &= \frac{1}{q-1}\log_2 \sum_{x\,\in\,X}\left(\frac{q(x)^q}{p(x)^{q-1}}\right) \\ &= \frac{1}{q-1}\log_2 \sum_{x\,\in\,X} q(x)\left(\frac{q(x)}{p(x)}\right)^{q-1} \\ &\geq \frac{1}{q-1}\sum_{x\,\in\,X} q(x)\log_2\left(\frac{q(x)}{p(x)}\right)^{q-1} \\ &= D_{KL}(q_X||p_X) \;\geq\; 0\,. \end{aligned} \tag{D.5}$$

Here in the first inequality I have applied Jensen's inequality to the concave function $\log_2 x$. I could alternatively apply Jensen's inequality to the function x^q which is convex for $q > 1$ and concave if $0 < q < 1$. In the second inequality I have used the result (D.4). The equality in (D.5) holds iff $p(x) = q(x)$. Note that in the limit $q \to 1$ the relative Rényi entropy reduces to the Kullback–Leibler divergence.

In passing we can observe that the relative entropy remains well-defined even for continuous distributions, and furthermore is invariant under parameter transformations. Since Jensen's inequality is valid also for generic average values the above Gibbs' inequality (D.5) (and similarly (D.2)) remains valid for continuous distributions and equals to zero iff $p_X = q_X$ *almost everywhere*.

D.3 Mutual information vs. relative entropy

While in Shannon's case we have the relation between mutual information and relative entropy given by the identity

$$\begin{aligned} I_1(X;Y) &= H(X) \;-\; H(X|Y) \;=\; H(Y) \;-\; H(Y|X) \\ &= D_{KL}(p_{(X,Y)}||p_X\cdot p_Y) \\ &= \sum_{x\,\in\,X,\;y\,\in\,Y} p(x,y)\log_2\frac{p(x,y)}{p(x)q(y)}\,, \end{aligned} \tag{D.6}$$

no similar relation exists for information measures of the order q ($q \neq 1$). Indeed, in such a case we have

$$\begin{aligned} I_q(X;Y) &= \mathcal{I}_q(X) - \mathcal{I}_q(X|Y) = \mathcal{I}_q(Y) - \mathcal{I}_q(Y|X) \\ &= \frac{1}{q-1}\log_2 \frac{\sum_{x\in X,\, y\in Y} p^q(x,y)}{\sum_{x\in X,\, y\in Y} q^q(x)p^q(y)} \\ &\neq I_q(p_{(X,Y)}||p_X \cdot p_Y) \\ &= \frac{1}{q-1}\log_2 \sum_{x\in X,\, y\in Y} \frac{p^q(x,y)}{[q(x)p(y)]^{q-1}}\,. \end{aligned} \tag{D.7}$$

The later fact has two important consequences: *a*) while $I_q(p_{(X,Y)}||p_X \cdot p_Y)$ is always non-negative due to Gibbs' inequality, $I_q(X;Y)$ can be negative and *b*) while $I_q(p_{(X,Y)}||p_X \cdot p_Y) = 0$ implies that $p_{(X,Y)} = p_X \cdot p_Y$ (i.e. X and Y are independent), $I_q(X;Y) = 0$ does not imply that X and Y are independent (though if they are independent $I_q(X;Y)$ is clearly zero). We might, however, observe that

$$\lim_{q\to 1} I_q(X;Y) = \lim_{q\to 1} I_q(p_{(X,Y)}||p_X \cdot p_Y) = D_{KL}(p_{(X,Y)}||p_X \cdot p_Y)\,. \tag{D.8}$$

Above dissimilarity between mutual information and relative entropy in the Rényi entropy case is reason why Rényi transfer entropy (cf. Chapter 2.3.2) is so different in comparison with its Shannon's counterpart.

D.4 Proof of Eq. (2.90)

In this proof I will loosely follow Refs. [146, 158]. Let me assume that a random D-dimension vector X has a finite covariance. The non-Gaussianness of X can be quantified via Shannon's relative entropy (i.e., Kullback–Leibler divergence) with respect to a Gaussian random vector $\mathcal{Z}_G^K$ with identical second moments[6]. In other words, I wish to evaluate a gain of information when replacing the Gaussian prior distribution with non-Gaussian posterior distribution. The ensuing relative entropy (measured in *nats*) can be written as

$$\begin{aligned} I_1(X||\mathcal{Z}_G^K) &\equiv I_1(q_X||p_{G,X}) = \sum_{x\in X} q(x)\log\frac{q(x)}{p_G(x)} \\ &= H(\mathcal{Z}_G^K) - H(X)\,. \end{aligned} \tag{D.9}$$

[6] It is implicitly assumed that first moments of both distributions are the same and equal to zero. In case when the mean of X is non-zero, one must work with correspondingly shifted Gaussian random vector.

Here

$$p_G(x) \;=\; (2\pi)^{-D/2}|K|^{-1/2}\exp\left(-\frac{1}{2}x_i K^{-1}_{ij}x_j\right), \tag{D.10}$$

and $H(\mathcal{Z}^K_G) = \frac{1}{2}\log\Big[(2\pi e)^D\,|K|\Big]$. So, by using the formula for entropy power (2.81) I get

$$\begin{aligned} N(X) \;\equiv\; N_1(X) \;&=\; \frac{1}{2\pi e}\exp\left(\frac{2}{D}H(X)\right) \\ &=\; \frac{1}{2\pi e}\exp\left(\frac{2}{D}H(\mathcal{Z}^K_G)\right)\exp\left(-\frac{2}{D}I_1(X||\mathcal{Z}^K_G)\right) \\ &\leq\; \frac{1}{2\pi e}\exp\left(\frac{2}{D}H(\mathcal{Z}^K_G)\right) \;=\; |K|^{1/D}\,. \end{aligned} \tag{D.11}$$

I can further use the fact that

$$\log(\det\mathbb{A}) \;=\; \mathrm{Tr}(\log\mathbb{A})\,, \tag{D.12}$$

which is certainly valid for any diagonalizable matrix $\mathbb{A}$, and more generally for *all matrices* since diagonalizable matrices are dense. With this I can write

$$\begin{aligned} \log|K|^{1/D} \;&=\; \sum_{i=1}^{D}\left[\frac{1}{D}\log(K)_{ii}\right] \\ &\leq\; \log\left[\mathrm{Tr}(K_{ij})/D\right] \;=\; \log\sigma^2\,, \end{aligned} \tag{D.13}$$

where σ^2 is the variance per component. The inequality follows from Jensen's inequality for the logarithm. Note that (D.13) directly implies that $|K|^{1/D} \leq \sigma^2$. The equality in (D.11) is saturated iff X is a Gaussian vector and the inequality in (D.13) is saturated iff the covariance matrix K_{ij} is proportional to the identity matrix. This completes the proof of Eq. (2.90).

Appendix E

Riesz–Thorin inequality

To prove the ITUR (2.64) I need to prove a particular variant of the Riesz–Thorin inequality [118, 268, 306] upon which my considerations are based. For this purpose I first state the Riesz convexity theorem.

Theorem E.5 (Riesz convexity theorem). *Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{C}^n$ and $\mathcal{L}$ be a linear operator such that $(\mathcal{L}\boldsymbol{x})_j = \sum_i a_{ij} x_j$. Let, in addition, $M_{\alpha\beta}$ be the least number "k" satisfying*

$$\|\mathcal{L}\boldsymbol{x}\|_{1/(1-\beta)} \leq k\|\boldsymbol{x}\|_{1/\alpha}\,,$$

where $\|\boldsymbol{x}\|_p = (\sum_i |x_i|^p)^{1/p}$. Then, $\log(M_{\alpha\beta})$ is convex in triangle $0 \leq \alpha; \beta \leq 1,\ \alpha + \beta \geq 1$.

The convexity triangle is depicted in Figure E.1. Detailed exposition of the proof can be found for example in [118].

Corollary E.1. *Let (α_1, β_1) and (α_2, β_2) be two points in the above convex triangle. If I define*

$$\alpha = \alpha_1 s + \alpha_2(1-s)\,,\quad \beta = \beta_1 s + \beta_2(1-s)\,;\qquad s \in [0,1]\,,$$

then clearly

$$\log(M_{\alpha\beta}) \leq s\log(M_{\alpha_1\beta_1}) + (1-s)\log(M_{\alpha_2\beta_2})\,,$$

or equivalently

$$M_{\alpha\beta} \leq M^s_{\alpha_1\beta_1} M^{(1-s)}_{\alpha_2\beta_2}\,.$$

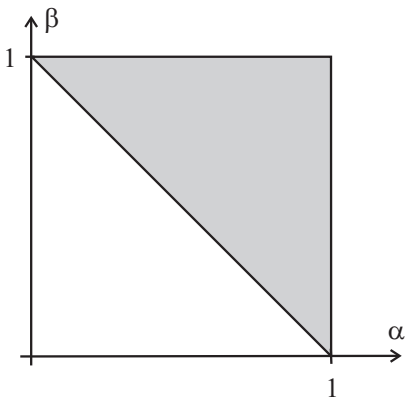


Fig. E.1: Riesz convexity triangle. Riesz's inequality in Theorem 1 holds only when α and β belong to the shaded region.

Theorem E.6 (Riesz–Thorin inequality). *Let* $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{C}^n$. *Suppose that* $(\mathcal{L}\boldsymbol{x})_j = \sum_i a_{ji} x_i$ *and that*

$$\sum_j |(\mathcal{L}\boldsymbol{x})_j|^2 \;\le\; \sum_j |x_j|^2 \,.$$

Then for $p \in [1,2]$ *and* $c \equiv \max_{i,j} |a_{ij}|$

$$\|\mathcal{L}\boldsymbol{x}\|_{p'} \;\le\; c^{(2-p)/p} \|\boldsymbol{x}\|_p \;=\; c^{1/p} c^{-1/p'} \|\boldsymbol{x}\|_p \quad\Leftrightarrow\quad c^{1/p'} \|\mathcal{L}\boldsymbol{x}\|_{p'} \;\le\; c^{1/p} \|\boldsymbol{x}\|_p \,,$$

holds. Here p *and* p' *are Hölder conjugates, i.e.,* $1/p + 1/p' = 1$.

Proof of Theorem E.6. I will use the notation $\alpha = 1/p$, $\beta = 1/q$ (and the Hölder conjugates $p' = p/(p-1)$, $q' = q/(q-1)$). Consider the line from $(\alpha_1, \beta_1) = (\frac{1}{2}, \frac{1}{2})$ to $(\alpha_2, \beta_2) = (1,1)$ in the (α, β) plane. This line lies entirely in the triangle of concavity (see Figure E.1). Let me now define

$$\alpha \;=\; \alpha_1 s + \alpha_2 (1-s) \;=\; s/2 + (1-s) \;=\; -s/2 + 1 \,,$$

implying $s = 2(1-\alpha)$, and I also define

$$\beta \;=\; \beta_1 s \;+\; \beta_2 (1-s) \;=\; -s/2 + 1 \,,$$

implying $\beta = \alpha$. Hence by Corollary E.1

$$M_{\alpha,\alpha} \;\le\; M^s_{\alpha_1\beta_1} M^{(1-s)}_{\alpha_2\beta_2} \;=\; M^{2(1-\alpha)}_{\frac{1}{2},\frac{1}{2}} \, M^{2\alpha-1}_{1,1} \,. \tag{E.1}$$

Note particularly that because $s \in [0,1]$ then $\alpha \in [\frac{1}{2}, 1]$ and $p \in [1,2]$.

To estimate the right hand side of (E.1) I use that $M_{\frac{1}{2},\frac{1}{2}} \leq 1$. This results from the very assumption of the theorem, namely that

$$\|\mathcal{L}\boldsymbol{x}\|_2^2 \;=\; \sum_j |(\mathcal{L}\boldsymbol{x})_j|^2 \;\leq\; \sum_j |x_j|^2 \;=\; \|\boldsymbol{x}\|_2^2 \,.$$

Hence, $M_{\frac{1}{2},\frac{1}{2}} \leq k = 1$. To find the estimate for M_{11} I employ that it represents the smallest k in the relation

$$\|\mathcal{L}\boldsymbol{x}\|_\infty \;\leq\; k\|\boldsymbol{x}\|_1 \,.$$

Thus

$$M_{11} \;=\; \max_{\boldsymbol{x}\neq 0} \frac{\|\mathcal{L}\boldsymbol{x}\|_\infty}{\|\boldsymbol{x}\|_1} \;=\; \max_{\boldsymbol{x}\neq 0} \frac{\max_j |(\mathcal{L}\boldsymbol{x})_j|}{\sum_i |x_i|} \;\leq\; \max_{i,j} |a_{ij}| \;\equiv\; c\,.$$

So, finally I can write that

$$M_{\alpha,\alpha} \;=\; M_{1/p,(1-1/p')} \;\leq\; c^{2\alpha-1} \;=\; c^{(2-p)/p} \;=\; c^{1/p}c^{-1/p'} \,,$$

which proves the theorem. □

E.1 Connection with Theorem 2.3

To establish the connection with the Theorem 2.3 let me assume that $\mathcal{X}$ is a discrete random variable with n different values, $\mathbb{P}_n$ is the probability space affiliated with $\mathcal{X}$ and $\mathcal{P} = \{p_1, \ldots, p_n\}$ is a sample probability distribution from $\mathbb{P}_n$. Normally the geometry of $\mathbb{P}_n$ is identified with the geometry of a simplex. For our purpose it is more interesting to embed $\mathbb{P}_n$ in a sphere. Because $\mathcal{P}$ is non–negative and summable to unity, it follows that the square–root likelihood $|x_i| \equiv \xi_i = \sqrt{p_i}$ exists for all $i = 1, \ldots, n$, and it satisfies the normalization condition

$$\sum_{i=1}^{n} (\xi_i)^2 = 1\,.$$

Hence $\boldsymbol{\xi}$ can be regarded as a unit vector in the Hilbert space $\mathcal{H} = \mathbb{R}^n$. Then the inner product

$$\cos\phi = \sum_{i=1}^{n} \xi_i^{(1)}\xi_i^{(2)} = 1 - \frac{1}{2}\sum_{i=1}^{n} \left(\xi_i^{(1)} - \xi_i^{(2)}\right)^2 \,, \tag{E.2}$$

defines the angle ϕ that can be interpreted as a distance between two probability distributions. More precisely, if $\mathcal{S}^{n-1}$ is the unit sphere in the n-dimensional Hilbert space, then ϕ is the spherical (or geodesic) distance between the points on $\mathcal{S}^{n-1}$ determined by $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$.

Now, let $\mathcal{P}^{(1)}$ and $\mathcal{P}^{(2)}$ denote a pair of probability distributions and $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$ the corresponding elements in Hilbert space. Because $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$ are non-negative, they are located only on the positive orthant of $\mathcal{S}^{n-1}$.

To proceed, I now set $p' = 2(1+t)$ and $p = 2(1+r)$ (remembering that $1/p + 1/p' = 1$). Then the Riesz–Thorin inequality reads (substituting $|x_i|$ for $\xi_i^{(1)}$)

$$\left(\sum_i (\xi_i^{(2)})^{p'}\right)^{1/p'} \;\leq\; c^{(2-p)/p} \left(\sum_i (\xi_i^{(1)})^{p}\right)^{1/p}, \tag{E.3}$$

which is equivalent to

$$\left(\sum_j (p_j^{(2)})^{(1+t)}\right)^{1/2(1+t)} \left(\sum_k (p_k^{(1)})^{(1+r)}\right)^{-1/2(1+r)} \;\leq\; c^{-r/(1+r)}\,.$$

I raise both sides to the power $2\,(1+t)\,/t$ and get

$$\left(\sum_j (p_j^{(2)})^{(1+t)}\right)^{1/t} \left(\sum_k (p_k^{(1)})^{(1+r)}\right)^{-(1+t)/t(1+r)}$$

$$\leq\; c^{-2r(1+t)/t(1+r)}\,. \tag{E.4}$$

The parameters are limited due to the condition $p \in [1,2]$ and $1/p + 1/p' = 1$ implying that

$$t \;=\; -r/(2r+1)\,. \tag{E.5}$$

This implies that $r \in [-1/2, 0]$ and $t \in [0, \infty)$. Combining (E.4) and (E.5) we get

$$\left(\sum_j (p_j^{(2)})^{(1+t)}\right)^{1/t} \left(\sum_k (p_k^{(1)})^{(1+r)}\right)^{1/r} \;\leq\; c^2\,. \tag{E.6}$$

By applying the negative binary logarithm on both sides of (E.6) I obtain the Theorem 2.3.

Appendix F

Beckner–Babenko theorem

To prove the Beckner–Babebko's theorem 2.4, I follow the exposition from Ref. [146]. I start with the (generalized) Young inequality that is instrumental in the proof. Since the actual proof of Young's inequality is rather involved I provide only its statement. The reader can find the proof together with further details, e.g., in Ref. [200].

Theorem F.7 (Young's theorem). *Let* $q, p, r > 0$ *represent Hölder triple, i.e.,*

$$\frac{1}{q} + \frac{1}{p} = 1 + \frac{1}{r},$$

and let $\mathcal{F} \in \ell^q(\mathbb{R}^D)$ *and* $\mathcal{G} \in \ell^p(\mathbb{R}^D)$ *are two non-negative functions, then*

$$\|\mathcal{F} * \mathcal{G}\|_r \geq C^D \|\mathcal{F}\|_q \|\mathcal{G}\|_p , \tag{F.1}$$

for $q, p, r \geq 1$ *and*

$$\|\mathcal{F} * \mathcal{G}\|_r \leq C^D \|\mathcal{F}\|_q \|\mathcal{G}\|_p , \tag{F.2}$$

for $q, p, r \leq 1$. *The constant* C *is*

$$C = C_p C_q / C_r \quad \text{with} \quad C_x^2 = \frac{|x|^{1/x}}{|x'|^{1/x'}} .$$

Here x *and* x' *are Hölder conjugates. Symbol* $*$ *denotes a convolution.*

Let us now observe that the following chain of reasonings holds

$$\|\mathcal{F} * \delta\|_r \geq C^D \|\mathcal{F}\|_q \|\delta\|_p = C^D \|\mathcal{F}\|_q V_R^{(p-1)/p} . \tag{F.3}$$

Here I have used the fact that for the δ function

$$\|\delta\|_p = \left[\int_{\mathbb{R}^D} d\boldsymbol{x}\, \delta^p(\boldsymbol{x}) \right]^{1/p} = \left[\int_{\mathbb{R}^D} d\boldsymbol{x}\, \delta(\boldsymbol{x}) \delta^{p-1}(0) \right]^{1/p} = V_R^{(p-1)/p} .$$

In the derivation I have utilized that

$$\delta(0) \;=\; \int_{\mathbb{R}^D} d\boldsymbol{x}\, e^{i\boldsymbol{p}\cdot\mathbf{0}} \;=\; V_R\,.$$

Subindex R indicates that the volume is regularized, i.e., I approximate the actual volume of $\mathbb{R}^D$ with a D-dimensional ball of the radius R, where R is arbitrarily large but fixed. At the end of calculations I send R to infinity. I should also stress that in (F.3) an implicit assumption was made that $q, p, r \geq 1$.

The norm $\|\mathcal{F} * \delta\|_r$ fulfills yet another inequality, namely

$$\begin{aligned}\|\mathcal{F} * \delta\|_r \;&=\; \left[\int_{\mathbb{R}^D} d\boldsymbol{x}\, \left(\int_{\mathbb{R}^D} d\boldsymbol{p}\, e^{-i\mathbf{p}\cdot\boldsymbol{x}}\hat{\mathcal{F}}(\boldsymbol{p})\right)^r\right]^{1/r} \\ &\leq\; \|\hat{\mathcal{F}}\|_n V_R^{1/n'+1/r}\,, \end{aligned} \tag{F.4}$$

where I have used the Hölder inequality

$$\begin{aligned}\int_{\mathbb{R}^D} d\boldsymbol{p}\, e^{-i\mathbf{p}\cdot\boldsymbol{x}}\hat{\mathcal{F}}(\boldsymbol{p}) \;&=\; \left|\int_{\mathbb{R}^D} d\boldsymbol{p}\, e^{-i\mathbf{p}\cdot\boldsymbol{x}}\hat{\mathcal{F}}(\boldsymbol{p})\right| \leq \|\hat{\mathcal{F}}\|_n \|e^{-i\boldsymbol{p}\cdot\boldsymbol{x}}\|_{n'} \\ &=\; \|\hat{\mathcal{F}}\|_n V_R^{1/n'}\,, \end{aligned} \tag{F.5}$$

with n and n' being Hölder's conjugates ($n \geq 1$).

Comparing (F.3) with (F.4) I get the inequality

$$\|\hat{\mathcal{F}}\|_n V_R^{1/n'+1/r} \;\geq\; C^D \|\mathcal{F}\|_q V_R^{(p-1)/p}\,. \tag{F.6}$$

The inequality is explicitly volume independent provided $1/n'+1/r+1/p = 1$, or equivalently, when $1/n' = 1/q - 2/r$. With this I can rewrite (F.6) as

$$\|\hat{\mathcal{F}}\|_n \;\geq\; C^D \|\mathcal{F}\|_q \;\geq\; C^D \|\mathcal{F}\|_{n'}\,. \tag{F.7}$$

The last inequality results from Hölder's inequality

$$\|\mathcal{F}\|_a \;\geq\; \|\mathcal{F}\|_b \qquad \text{when} \qquad a \leq b\,. \tag{F.8}$$

In fact, in the limit $r \to \infty$ the last inequality in (F.7) is saturated and $C \overset{r\to\infty}{\to} 1$. Consequently I get the so-called *Hausdorff–Young inequality* in the form

$$\|\hat{\mathcal{F}}\|_n \;\geq\; \|\mathcal{F}\|_{n'}\,. \tag{F.9}$$

This inequality holds, of course, only when $q \geq n'$ (cf. equation (F.8)), i.e., when $n \geq q/(q-1)$. Since $q \geq 1$ one has that $n \in [1, 2]$. Should I have started in the derivation with $\hat{\mathcal{F}}$ instead of $\mathcal{F}$, I would have obtain the reverse inequality

$$\|\mathcal{F}\|_n \;\geq\; \|\hat{\mathcal{F}}\|_{n'}\,. \tag{F.10}$$

Inequalities, (F.9) and (F.10) are known as classical Hausdorff–Young inequalities [118]. Note that in the spacial case when $n = 2$ also $n' = 2$ and equations (F.9) - (F.10) together imply equality

$$\|\mathcal{F}\|_2 \;=\; \|\hat{\mathcal{F}}\|_2\,. \tag{F.11}$$

This is the familiar Plancherel (or Riesz–Fischer or Parseval) equality [118].

It should be noted that the Beckner–Babenko inequality as stated in Theorem 2.4 improves upon the Hausdorff–Young inequalities. This is because $C_x \leq 1$ for $x \in [1, 2]$, see Fig. F.1. The Beckner–Babenko inequality

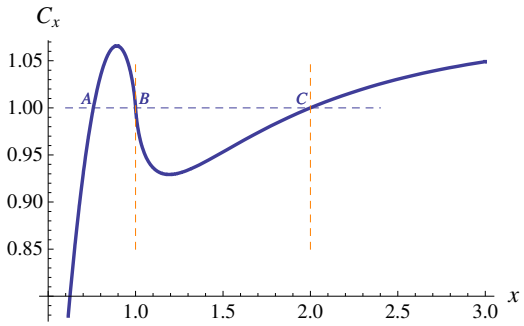


Fig. F.1: Dependence of the constant C_x on the Hölder parameter x. When x is between points B and C, i.e., when $x \in [1, 2]$ than $C_x \leq 1$. For $x \leq A$ is C_x also smaller than 1 but such x are excluded by the fact that x must be ≥ 1.

now follows easily from Young's inequality. Indeed, assume that there exists a (possibly p-dependent) constant $k(p) \leq 1$, such that

$$k(p)\|\mathcal{F}\|_p \;\geq\; \|\hat{\mathcal{F}}\|_{p'} \qquad \text{and} \qquad k(p)\|\hat{\mathcal{F}}\|_p \;\geq\; \|\mathcal{F}\|_{p'}\,. \tag{F.12}$$

The constant $k(p)$ can be easily found by writing

$$\begin{aligned} k(r)\|\mathcal{F} * \mathcal{G}\|_r \;&\geq\; \|\hat{\mathcal{F}}\hat{\mathcal{G}}\|_{r'} \;\geq\; \|\hat{\mathcal{F}}\|_{q'}\|\hat{\mathcal{G}}\|_{p'} \\ &\geq\; [k(q')]^{-1}\|\mathcal{F}\|_q[k(p')]^{-1}\|\mathcal{G}\|_p\,, \end{aligned} \tag{F.13}$$

which gives

$$\|\mathcal{F} * \mathcal{G}\|_r \;\geq\; [k(r)]^{-1}[k(q')]^{-1}[k(p')]^{-1}\|\mathcal{F}\|_q\|\mathcal{G}\|_p\,. \tag{F.14}$$

The middle inequality in (F.13) is the Hölder inequality that is valid for $1/r' = 1/p' + 1/q'$ (i.e., for $1/p + 1/q = 1/r + 1$). Comparison of (F.14) with (F.1) gives the equation

$$\begin{aligned} [k(r)]^{-1}[k(q')]^{-1}[k(p')]^{-1} \;&=\; C^D \;=\; [C_qC_p/C_r]^D \\ &=\; [1/C_{q'}C_{p'}C_r]^D\,. \end{aligned} \tag{F.15}$$

This is clearly solved with $k(p) = C_p^D$. By choosing $p \in [1, 2]$ I get the statement of the Beckner–Babenko theorem.

Chapter 3

Superstatistics

In this chapter I consider yet another important concept which gives rise to generalized statistics, namely *superstatiscs*. In contrast to previous approaches, where generalized statistics enters via non-Gibbsian (or non-Shannonian) entropic measures, the superstatistics paradigm is based on a superposition of several dynamics and their respective statistics. In particular, the entropy is not a pivotal concept in superstatistics. Among a rich palette of applications I discuss applications in finance in option pricing and in quantum mechanics of a relativistic particle. A remarkable connection of superstatistics with a doubly-special relativity is also briefly discussed.

3.1 Old wine in new bottles

Complex dynamical systems often demonstrate patterns of behavior that can be regarded as a superposition of several dynamics operating on different spatio-temporal scales. The corresponding effects then produce a superposition of respective statistics, or in short *superstatistics* [23, 24].

Christian Beck, one of the founders of the superstatistics concept vividly describes the situation as follows [24]:

"Many non-equilibrium systems actually exhibit spatial or temporal temperature fluctuations on a rather large scale. Think, for example, of the weather: It is unlikely that the temperature in London, New York, and Firenze is the same at the same time. There are spatio-temporal temperature fluctuations on a rather large scale, though locally equilibrium statistical mechanics with a given fixed temperature is certainly valid. A traveler who frequently travels between the three cities

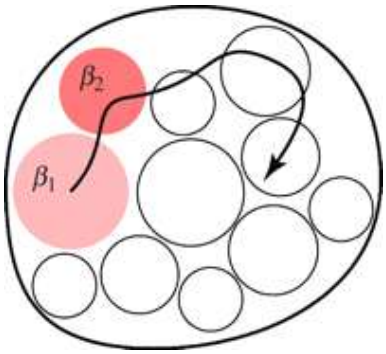


Fig. 3.1: A Brownian particle moving through spatio-temporal inhomogeneous environment with temperature fluctuations on a large scale.

sees a "mixture" of canonical ensembles corresponding to different local temperatures. Such type of macroscopic inhomogeneities of an intensive parameter occur not only for the weather but for many other driven nonequilibrium systems as well."

The important point in the above description of the superstatistics is the existence of two well separated time scales, say τ and T. The local (or relaxation) time scale τ is the relevant time scale on which the particle equilibrates with its local environment. The large time scale $T \gg \tau$, is the correlation time, i.e., the relevant time scale over which β appreciably changes (cf. Fig. 3.1).

Previously described situation can be also nicely illustrated with a Brownian particle suspended in thermally inhomogeneous fluid environment, see Fig. 3.2. In this case the probability to find a Brownian particle to have velocity $\boldsymbol{v}$ in the interval $(\boldsymbol{v}, \boldsymbol{v} + d\boldsymbol{v})$ is

$$p(\boldsymbol{v})d\boldsymbol{v} \;=\; \int p(\boldsymbol{v}|\beta)\, f(\beta) d\beta d\boldsymbol{v} \;\propto\; \int e^{-\beta m \boldsymbol{v}^2/2}\, f(\beta) d\beta d\boldsymbol{v}\,. \qquad (3.1)$$

The latter is the usual marginalization prescription with β serving as a nuisance parameter, $p(\boldsymbol{v}|\beta)$ as a conditional PDF (which is Maxwell–Boltzmann by assumption) and $p(\boldsymbol{v})$ as a marginal PDF. The marginal PDF thus describes the long-term behavior of the particle for times $T \gg \tau$. Let me emphasize that the prescription (3.1) have sense only when $p(\boldsymbol{v}|\beta)$ is operationally well defined, which in turn, is true only when the two relevant scales are well separated. According to terminology used in classical and

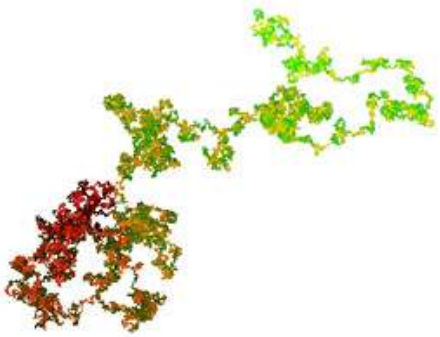


Fig. 3.2: Example of Brownian particle in thermally inhomogeneous fluid environment. A different color denotes a different local equilibrium temperature $1/\beta$.

quantum mechanics, the above classifies as *adiabatic process* since gradually changing conditions (in our case β) allow the system to adapt its configuration (in our case the conditional distribution $p(\boldsymbol{v}|\beta)$).

Let me finally stress that the intensive parameter in question not necessarily need to be (inverse) temperature β. In the literature one can find also other pertinent parameters, for instance, energy dissipation rate (turbulent flow in Kolmogorov theory), volatility (econophysics), einbein (quantum relativistic particle), etc.

3.1.0.1 *Toy-model motivation — Tsallis' thermostatistics*

One of the simplest, and at the same time conceptually very important is the reconstruction of Tsallis-type distributions out of superstatistics. This was first performed in Ref. [24], and the corresponding path-integral derivation was originally presented in Ref. [148]. To this end, I can consider Tsallis' probability density and set $q = \epsilon + 1$ then

$$\rho_{_T}(\beta) = \frac{1}{Z_q}\left[1 + (q-1)\beta H\right]^{1/(1-q)} = \frac{1}{Z_q \Gamma(1/\varepsilon)} \int_0^\infty \frac{dt}{t}\, t^{1/\varepsilon} e^{-t} e^{-\varepsilon\beta t H}$$

$$= \int_0^\infty dv\, g\left(v; \frac{1}{\varepsilon\beta}, \frac{1}{\varepsilon}\right) \rho_{_G}(v)\,, \tag{3.2}$$

where $\rho_{_G}(H, v) \equiv p(H|v)$ represents the the Gibbsian PDF, namely

$$\rho_{_G}(v) = \frac{e^{-vH}}{Z_{_G}}\,, \tag{3.3}$$

and the smearing PDF

$$g(x;\mu,\nu) \;=\; \frac{1}{\Gamma(\nu)}\;\mu^{\nu}x^{\nu-1}e^{-\mu x}, \tag{3.4}$$

is the so-called *Gamma* distribution [85]. The corresponding first two momenta read [85]:

$$\langle x\rangle \;=\; \nu/\mu\,, \qquad \langle x^2\rangle \;=\; \frac{\nu(\nu+1)}{\mu^2}\,. \tag{3.5}$$

The form (3.2) allows to reinterpret Tsallis's distribution as Gibbs' distribution with the inverse temperature v weighted with Gamma PDF. In this picture the distribution parameters have a particularly clear meaning. With the help of (3.5) I have $\langle v\rangle = \beta$ and $\langle v^2\rangle/\langle v\rangle^2 = q$. In addition, because $\lim_{q\to 1_+} g(x;\mu,\nu) = \delta(x-\nu/\mu)$ I get that $\rho_{_T}(\beta) \to e^{-\beta_{_G} H}/Z_{_G}$ with $\beta_{_G} = \beta|_{q\to 1}$ representing the usual inverse equilibrium temperature. This again reconfirm the result of Chapter 1, i.e. that Tsallis statistics goes over into Gibbs–Boltzmann statistics in the $q \to 1_+$ limit. Once more, I should stress that the above picture has sense only when the associated relaxation and correlation time scales are well separated.

Important property of Gamma distribution is its *infinite divisibility* (for technical exposition of infinite divisibility see Appendix B.6). The infinite divisibility is, in fact, a key property allowing to formulate Tsallis thermostatistics in the path-integral language, cf. Refs. [148, 179]. I will return back to this point in Section 3.2.

3.1.1 *Universality Classes*

While in principle any smearing PDF $f(\beta)$ is possible in the superstatistics approach, in practice one usually observes only few relevant PDF's. These are the *Gamma*, *inverse Gamma* and *log-normal* distribution. In other words, in complex systems with (two-)time-scale separation one usually observes three physically relevant "universality classes" [25, 117]: a) Gamma-superstatistics (leading to Tsallis thermostatistics), b) inverse Gamma-superstatistics and c) log-normal-superstatistics. The reason for these classes can be traced, according to Beck–Gell-Mann *et al.* [25, 117], to three typical phenomenological situations that could be realistically responsible for emergence of the random variable β.

The basic assumption involved is that β results from a collective (or cumulative) effect of a large number of independent microscopic random variables, say $X_i, i = 1,\ldots,N$, which are present in the system. If this

is so, then the relevant variable is $\hat{S}_N = [\sum_{i=1}^N X_i]/\sqrt{N}$ (assuming, for simplicity, zero mean). Presuming validity of the classical CLT, $\hat{S}_N$ will approach in the large-N limit a Gaussian random variable, say $\bar{S}$. Since, β is by definition positive, it can be obtained by squaring $\bar{S}$, so that $\beta \propto \bar{S}^2$. The distribution for β is thus

$$\begin{aligned} f(\beta) &= \int_{\mathbb{R}} dx\, \delta(cx^2 - \beta)\, \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/(2\sigma^2)} \\ &= \int_{\mathbb{R}} dx \left(\frac{\delta(x - \sqrt{\beta/c})}{2\sqrt{c\beta}} + \frac{\delta(x + \sqrt{\beta/c})}{2\sqrt{c\beta}} \right) \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/(2\sigma^2)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2 c\beta}}\, e^{-\beta/(2c\sigma^2)}. \end{aligned} \tag{3.6}$$

Which is nothing but the Gamma distribution with $2c\sigma^2$ representing the scale and $c > 0$ being the proportionality constant between random variables β and $\bar{S}^2$. The full Gamma distribution appears when there are more than just 1 (say k) Gaussian random variables $\bar{S}_1, \ldots, \bar{S}_k$ corresponding to relevant degrees of freedom in the system that contribute to β. In this case I indeed get (assuming, for simplicity, equal variances)

$$\begin{aligned} f(\beta) &= \int_{\mathbb{R}^k} d\mathbf{x}\, \delta(c\mathbf{x}^2 - \beta)\, \frac{1}{(2\pi\sigma^2)^{k/2}}\, e^{-\mathbf{x}^2/(2\sigma^2)} \\ &= \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty dr\, r^{k-1}\, \frac{\delta(r - \sqrt{\beta/c})}{2\sqrt{c\beta}}\, \frac{e^{-r^2/(2\sigma^2)}}{(2\pi\sigma^2)^{k/2}} \\ &= \frac{\beta^{k/2-1}}{(2\sigma^2 c)^{k/2}\Gamma(k/2)}\, e^{-\beta/(2c\sigma^2)}\,. \end{aligned} \tag{3.7}$$

Which is the defining distribution of the *Gamma-superstatistics* [25].

The *inverse Gamma-superstatistics* is obtained by assuming that $\beta \propto 1/\bar{S}^2$ or, in case of more (say k) relevant degrees of freedom, $\beta \propto 1/(\sum_1^k \bar{S}_i^2)$. This situation is clearly more suited for temperature rather than inverse temperature. In this case the smearing distribution reads

$$\begin{aligned} f(\beta) &= \int_{\mathbb{R}^k} d\mathbf{x}\, \delta\left(\frac{c}{\mathbf{x}^2} - \beta\right)\, \frac{1}{(2\pi\sigma^2)^{k/2}}\, e^{-\mathbf{x}^2/(2\sigma^2)} \\ &= \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty dr\, r^{k+2}\, \frac{\delta(r - \sqrt{c/\beta})}{2c}\, \frac{e^{-r^2/(2\sigma^2)}}{(2\pi\sigma^2)^{k/2}} \\ &= \frac{\beta^{-k/2-1}}{(2\sigma^2/c)^{k/2}\Gamma(k/2)}\, e^{-c/(2\beta\sigma^2)}\,. \end{aligned} \tag{3.8}$$

Finally, the *log-normal-superstatistics* is based on the idea that β results from a random multiplicative process rather than random additive process. This is, for instance, the situation known from multifractal processes [80, 296]. By using the multiplicative version of CLT (which ensures that the product of many independent, equally distributed positive random variables approaches a log-normal distribution) I can write the smearing distribution in the form

$$\begin{aligned} f(\beta) &= \int_{\mathbb{R}^k} d\mathbf{x}\, \frac{\delta\Big(c\prod_{i=1}^k x_i - \beta\Big)}{(2\pi\sigma^2)^{k/2}\prod_{i=1}^k x_i}\, \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^k \log x_i - k\mu\right)^2\right] \\ &= \frac{c}{(2\pi\sigma^2)^{k/2}\beta}\, \exp\left[-\frac{1}{2\sigma^2}\left(\log\beta/c - k\mu\right)^2\right]. \end{aligned} \tag{3.9}$$

Here $\mu = \langle \log x_i\rangle$ and $\sigma^2 = \langle(\log x_i - \mu)^2\rangle$. For simplicity I have assumed identical means and variances, i.e., μ and σ are i-independent.

The above *universality* of the classes as presented by Beck *at al.* is thus a consequence of the *classical* CLT and positivity of β. It might be therefore expected that the above three types of superstatistics will be valid for a wide class of (fast-time scale) stochastic environments.

Though, in principle one can envisage more complicated situations, it is believed that most experimentally relevant cases fall into one of the above three universality classes (or simple combinations of them). Gamma-superstatistics seems to be relevant for cosmic ray statistics [21], candidate systems for inverse gamma-superstatistics are systems exhibiting velocity distributions with exponential tails [276, 310], cancer survival statistics [58] or quantum relativistic particles [149, 159, 160], and log-normal superstatistics has been found for Lagrangian [23, 261] and Eulerian [22] turbulence. Existence of transitions between various superstatistics universality classes was demonstrated in the context of financial time series in Ref. [155]. In addition, in Ref. [148] it was shown that gamma and inverse gamma-superstatistics naturally appear in the framework of so-called superstatistics path integrals (cf. also Chapter 3.2).

3.2 Superstatistics path integral

The path integral (PI) has been used in quantum physics since the revolutionary work of Feynman [87, 91], although the basic observation goes back to Dirac [71, 73] who appreciated the rôle of the Lagrangian in

short-time evolution of the wave function, and even suggested the time-slicing procedure for finite, i.e., non-infinitesimal, time lags. Since then the PI approach has yielded invaluable insights into the structure of quantum theory [92, 282] and provided a viable alternative to the traditional operator-formalism-based canonical quantization. During the second half of the 20th century, the PI became a standard tool in quantum field theory [70, 178, 258] and statistical physics [335], often providing the easiest route to derivation of perturbative expansions and serving as an excellent framework for (both numerical and analytical) non-perturbative analysis [179]. Because of its overall space-time viewpoint, PI has also often helped to shed a new light on such fundamental issues as, quantization of gravity [6, 7, 70, 123] or alternative interpretations of quantum theory [39, 40, 107, 108].

Feynman's PI has its counterpart in pure mathematics, namely, in the theory of continuous-time stochastic processes [266]. There the concept of integration over a space of continuous functions (so-called fluctuating paths or sample paths) had been introduced by Wiener [330] already in 1920's in order to represent and quantify the Brownian motion. Interestingly enough, this so-called Wiener integral (or integral with respect to Wiener measure) was formulated 2 years before the discovery of the Schrödinger equation and 25 years before Feynman's PI formulation. I present some necessary foundations of both Feynman's and Wiener's PI in Appendix I.

In works [148, 149, 159, 160] it was shown that PI when coupled with the superstatistics paradigm can have interesting and unexpected implications in relativistic quantum physics. To understand what was at stake, I need to introduce the concept of the so-called *superstatistics path integral*.

I start with the key observation that when a conditional probability density function[1] can be formulated in terms of a PI, then it inevitably satisfies the Chapman–Kolmogorov semigroup equation (CKE) for Marcovian process [92, 282]

$$P(y,t''|x,t) \;=\; \int_{-\infty}^{\infty} dz\, P(y,t''|z,t')P(z,t'|x,t)\,, \tag{3.10}$$

with t' being any time between t'' and t. Conversely, any transition probability satisfying CKE possesses a PI representation [92]. In physics one

[1] *Conditonal probability* density function with the boundary condition $P(y,t|x,t) = \delta(y-x)$ is called *transitional probability* density function. In the context of PI the conditional probabilities involved are always implicitly assumed to be transitional probabilities, unless specified otherwise.

often encounters probabilities formulated as a superposition of PI's, namely

$$\wp(x',t'|x,t) = \int_0^\infty d\zeta\, \omega(\zeta,T) \int_{x(t)=x}^{x(t')=x'} \mathcal{D}x\, \mathcal{D}p\; e^{\int_t^{t'} d\tau [ip\dot{x} - \zeta H(p,x)]}\,. \quad (3.11)$$

Here $\omega(\zeta,T)$ with $T = t'-t$ is a normalized PDF defined on $\mathbb{R}^+\times\,\mathbb{R}^+$. The form (3.11) is quite ubiquitous and it often appears in non-perturbative approximations to statistical partition functions, in polymer physics, in financial markets, in systems with reparametrization invariance, etc. The corresponding random variable ζ is then associated with the inverse temperature, coupling constant, volatility, vielbein, etc. For some explicit examples see e.g. Ref. [179] and citations therein.

The most general class of distributions $\omega(\zeta,T)$ on $\mathbb{R}^+\times\,\mathbb{R}^+$ for which the superposition of Markovian processes remain Markovian, i.e., when also $\wp(x',t'|x,t)$ itself satisfies the CKE (3.10), was found in Ref. [148]. The key in the proof is to note that in order to have (3.10) satisfied by $\wp$, the rescaled PDF $w(\zeta,T) \equiv \omega(\zeta/T,T)/T$ should satisfy the integral equation

$$w(\zeta,t_1+t_2) \;=\; \int_0^\zeta d\zeta'\, w(\zeta',t_1) w(\zeta-\zeta',t_2)\,. \quad (3.12)$$

At this stage it is convenient to perform the (one-sided) Laplace transform of $w(\zeta,t)$, i.e.

$$\tilde{w}(p_{_\zeta},T) \;=\; \int_0^\infty d\zeta\, e^{-p_{_\zeta}\zeta} w(\zeta,T)\,. \quad (3.13)$$

In terms of $\tilde{w}$ one can rephrase Eq. (3.12) as a simple functional equation

$$\tilde{w}(p_{_\zeta},t_1+t_2) \;=\; \tilde{w}(p_{_\zeta},t_1)\tilde{w}(p_{_\zeta},t_2)\,, \quad (3.14)$$

with $t_1,t_2 \in \mathbb{R}^+$. By assuming continuity in T, the functional equation (3.14) has the generic solution $\tilde{w}(p_{_\zeta},T) = \{\tilde{w}(p_{_\zeta},1)\}^T$. From this we see that the distribution of ζ at T is completely determined by the distribution of ζ at $T=1$. In addition, from (3.14) it can be also deduced that $\tilde{w}(p_{_\zeta},1) = \{\tilde{w}(p_{_\zeta},1/n)\}^n$ for any $n \in \mathbb{N}^+$, and hence $w(\zeta,1)$ is infinitely divisible (see Appendix B.6). The Lévy–Khinchine theorem [34, 84] (see also Appendix B.6) then ensures that $\log\tilde{w}(p_{_\zeta},T) \equiv -TF(p_{_\zeta})$ must have the generic form

$$\log\tilde{w}(p_{_\zeta},T) \;=\; -T\left[\alpha p_{_\zeta} + \int_0^\infty (1-e^{-p_{_\zeta}u})\mu(du)\right], \quad (3.15)$$

where $\alpha \geq 0$ is a drift constant and μ is some non-negative measure on $(0,\infty)$ satisfying $\int_{\mathbb{R}^+} \min(1,u)\mu(du) < \infty$. The measure μ is called the Lévy

measure of the random variable ζ. Finally, the Laplace inverse of $\tilde{w}(p_\zeta, T)$ yields $\omega(\zeta, T)$. In passing we note that the RHS of (3.15) is basically the cumulant-generating function[2].

Once $\omega(\zeta, T)$ is found, then $\wp(x', t'|x, t)$ possesses a PI representation on its own. What is the form of the new Hamiltonian $\mathcal{H}(p, x)$? To answer this question I rewrite (3.11) in Dirac's notation as

$$\begin{aligned} \wp(x', t'|x, t) &= \langle x'| \int_0^\infty d\zeta\, w(\zeta, T) e^{-\zeta \hat{H}} |x\rangle = \langle x'| \{\tilde{w}(\hat{H}, 1)\}^T |x\rangle \\ &= \langle x'| e^{-TF(\hat{H})} |x\rangle\,. \end{aligned} \tag{3.16}$$

Hence, the identification $\mathcal{H}(p, x) = F(H(p, x))$ can be made. Corresponding generalization to higher dimensions is also straightforward [148].

At this stage one might worry about the operator-ordering issue. For my purpose it suffices to note that when H is x-independent, the former relation between $\mathcal{H}$ and H is exact. In more general situations the Weyl ordering is a natural choice because in this case the required mid-point rule follows automatically and one does not need to invoke the gauge invariance [20, 149]. Further details related to the ordering issue can be found in Ref. [149]. Transitional probability (3.11) with w given by (3.13) and (3.15) has been introduced in Ref. [148] and due to its inherent PI nature dubbed as *superstatistics path integral*.

Superstatistics PI permits the calculation of transitional probabilities for number of doubly stochastic processes in a novel way. Some calculations along those lines were presented in Refs. [148, 149, 159, 160]. Particularly important class of superstatistics PI's is related to the choice of the Lévy–Khinchine function $F(p_\zeta) = a\sqrt{p_\zeta + b}$, with a and b being real constants. Such a choice leads to PI's for relativistic and doubly relativistic particle systems, which can be written as a superstatistics average over *nonrelativistic* single-particle paths. In the following two subsections I will briefly discuss these two cases.

[2]At this point I should stress that in Chapter I, I used for infinitely divisible distribution the Lévy–Khintchine representation of *characteristic functions*. Since in the present context of superstatistics PIs the Laplace (rather than Fourier) transform is a more natural tool, the Lévy–Khintchine representation of *moment-generating functions* is more pertinent.

3.3 Emergent special relativity

The Feynman transition amplitude for a relativistic particle or better its Euclidean version — transition probability, naturally fits into the structure of superstatistics PI discussed above.

To see this, we might note that when in (3.15) I chose $\alpha = 0$ and $\mu(du) = 1/(2\sqrt{\pi}u^{3/2})du$ then the Lévy–Khinchine function $F(p_{\varsigma}) = \sqrt{p_{\varsigma}}$. In Refs. [148, 149] it was shown that with such $F(p_{\varsigma})$ one can rewrite the well known Newton–Wigner (Euclidean) propagator [237] for a Klein–Gordon (i.e., spinless relativistic) particle, i.e.

$$\wp(\mathbf{x}', t'|\mathbf{x}, t) \;=\; \int_{\mathbf{x}(0)=\mathbf{x}}^{\mathbf{x}(T)=\mathbf{x}'} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, e^{\int_0^T d\tau \left[i\mathbf{p}\cdot\dot{\mathbf{x}} \,-\, c\sqrt{\mathbf{p}^2+m^2c^2}\right]}\,, \qquad (3.17)$$

as a superposition of *non-relativistic* free-particle PI's, namely

$$\begin{aligned}&\wp(\mathbf{x}', t'|\mathbf{x}, t)\\ &\quad= \int_0^{\infty} dv\, \omega(v,T) \int_{\mathbf{x}(0)=\mathbf{x}}^{\mathbf{x}(T)=\mathbf{x}'} \mathcal{D}\mathbf{x}\,\mathcal{D}\mathbf{p}\, e^{\int_0^T d\tau [i\mathbf{p}\cdot\dot{\mathbf{x}} \,-\, v(\mathbf{p}^2c^2+m^2c^4)]}\,, \qquad (3.18)\end{aligned}$$

with $\omega(v,T)$ being the Weibull distribution of order 1, i.e., $\omega(v,T) \equiv W(v,1,T)$ where Weibull's PDF of order a is defined as [85]

$$W(v,a,T) \;=\; \frac{a\exp\left(-a^2T/4v\right)}{2\sqrt{\pi}\sqrt{v^3/T}}\,. \qquad (3.19)$$

It should be noted that neither (3.17) nor (3.18) are genuine propagators for a wave equation of a relativistic spinless particle — Klein–Gordon equation. As stressed first by Stückelberg [298, 299] (see also Refs. [90, 329]), the true relativistic propagator must include also the negative energy spectrum, reflecting the existence of antiparticles. In Refs. [148, 149] it has been proved that this can be rectified provided one rephrases the Klein–Gordon equation in the so-called Feshbach–Villars (FV) representation [86, 149]

$$i\partial_t\Psi \;=\; \hat{H}_{\rm FV}(\mathbf{p})\Psi \quad \text{with} \quad \hat{H}_{\rm FV}(\mathbf{p}) \;=\; (\sigma_3 + i\sigma_2)\frac{\hat{\mathbf{p}}^2}{2m} \;+\; \sigma_3 mc^2\,. \quad (3.20)$$

Here Ψ is a two component wave function where the respective components are related by opposite parity.

The Schrödinger equation-like structure of the FV equation allows to set up the PI representation of the propagator, which in this case is 2×2 matrix. In Ref. [149] it has been shown that in order to handle the full PI representation of the Klein–Gordon particle it suffices to discuss the PI relation (3.18) alone. This is because the latter serves as a building block

of the matrix structure of the FV propagator from which the Klein–Gordon propagator can be reconstructed [149].

By changing in Eq. (3.18) the variable vc^2 to $1/2\tilde{m}$, the RHS of (3.18) can be rewritten in the form

$$\int_{\mathbf{x}(0)=\mathbf{x}}^{\mathbf{x}(T)=\mathbf{x}'} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, e^{\int_0^T d\tau \left[i\mathbf{p}\cdot\dot{\mathbf{x}} - c\sqrt{\mathbf{p}^2+m^2c^2}\right]}$$

$$= \int_0^\infty d\tilde{m}\, f_{\frac{1}{2}}\left(\tilde{m}, Tc^2, Tc^2m^2\right) \int_{\mathbf{x}(0)=\mathbf{x}}^{\mathbf{x}(T)=\mathbf{x}'} \mathcal{D}\mathbf{x}\,\mathcal{D}\mathbf{p}\, e^{\int_0^T d\tau \left[i\mathbf{p}\cdot\dot{\mathbf{x}} - \frac{\mathbf{p}^2}{2\tilde{m}} - mc^2\right]}. \quad (3.21)$$

Here the smearing PDF

$$f_p(z,a,b) \;=\; \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})}\, z^{p-1}\, e^{-(az+b/z)/2}\,, \quad (3.22)$$

is the generalized inverse Gaussian distribution [85](K_p is the modified Bessel function of the second kind with index p). The LHS of (3.21) represents the PI for the free spinless relativistic particle in the Newton–Wigner representation [237]. The full Klein–Gordon (KG) kernel which also contains the negative-energy spectrum can be obtained from (3.22) by considering the Feshbach–Villars representation of the KG equation and making the substitution [149]

$$f_{\frac{1}{2}}\left(\tilde{m}, Tc^2, Tc^2m^2\right) \mapsto \frac{1+\mathrm{sgn}(T)\sigma_3}{2}\, f_{\frac{1}{2}}\left(\tilde{m}, |T|c^2, |T|c^2m^2\right). \quad (3.23)$$

The matrix nature of the smearing distribution (σ_3 is the Pauli matrix) naturally includes the Feynman–Stuckelberg causal boundary condition and thus treats both particles and antiparticles in a symmetric way [86, 149].

The structure of (3.21) allows to interpret $\tilde{m}$ as a Galilean-invariant Newton-like mass which takes on continuous values distributed according to $f_{\frac{1}{2}}\left(\tilde{m}, Tc^2, Tc^2m^2\right)$ with $\langle\tilde{m}\rangle = m + 1/Tc^2$ and $\mathrm{var}(\tilde{m}) = m/Tc^2 + 2/T^2c^4$. Fluctuations of the Newtonian mass can be then depicted as originating, e.g., from particle's evolution in an inhomogeneous or granular medium. Granularity, as known, for instance, from solid-state systems, typically leads to corrections in the local dispersion relation [9] and hence to alterations in the local *effective mass*. The following picture thus emerges: on a short-distance scale, a non-relativistic particle can be envisaged as propagating via classical Brownian motion through a single grain with a local mass $\tilde{m}$. This fast-time process has a time scale $\sim 1/\tilde{m}c^2$. An averaged value of the local time scale represents a transient temporal scale $\langle 1/\tilde{m}c^2\rangle = 1/mc^2$ which coincides with particle's Compton time T_C — i.e., the time for light

to cross the particle's Compton wavelength. At time scales much longer than T_C (large-distance scale), the probability that the particle encounters a grain which endows it with a mass $\tilde{m}$ is $f_{\frac{1}{2}}\big(\tilde{m}, Tc^2, Tc^2m^2\big)$. As a result one may view a single-particle relativistic theory as a single-particle non-relativistic theory where the particle's Newtonian mass $\tilde{m}$ represents a fluctuating parameter which approaches on average the Einstein rest mass m in the large T limit. I stress that T should be understood as the *observation time*, a time after which the observation (position measurement) is made. In particular, during the period T the system remains unperturbed. One can thus justly expect that in the long run all mass fluctuations will be washed out and only a sharp time-independent effective mass will be perceived. From the form of $\langle\tilde{m}\rangle$ it can be seen that the time scale at which this happens is $T \sim 1/mc^2$, i.e. the Compton time T_C. One might also observe that by coarse-graining the velocity over the correlation time $T_{_C}$ one gets

$$\langle|\mathbf{v}|\rangle_{_{T_C}} \;=\; \frac{\langle|\mathbf{p}|\rangle}{\langle\tilde{m}\rangle}\bigg|_{T_C} \;=\; c\,. \tag{3.24}$$

So, on a short-distance scale of order λ_C the spinless relativistic particle propagates with an averaged velocity, which equals to the speed of light c. But if one checks the particle's position at widely separated intervals (much larger than λ_C), then many directional reversals along a typical PI trajectory will take place, and the particle's net velocity will be then less than c — as it should be for a massive particle. The particle then acquires a sharp mass equal to Einstein's mass, and the process (not being hindered by fluctuating masses) is purely Brownian. This conclusion is in line with the Feynman checkerboard (FC) picture [137, 160] to which it reduces in the case of $(1+1)$ dimensional relativistic Dirac particle.

To appreciate the latter connection, I briefly mention the essence of FC approach (see also Fig. K.1). Feynman published his idea about FC only in the form of a short note in his book with Hibbs [92]. Apparently it was Schweber who discovered, while going through Feynman's notes, that the more complete formulation of the FC is dated as early as 1946 (for details see Ref. [284]). The chessboard formulation was also independently found by Riazanov [267]. In its essence, the FC is a way how to calculate (causal) Green's function of a massive Dirac fermion in $1+1$ dimensions. There the (massive!) electron propagates at the speed of light c. At random Poisson distributed moments of time its helicity and direction of motion are simultaneously changed and the average path traversed between these

moments of time is equal to λ_C. Feynman hoped that with FC he could explain particle's spin as a result of the space-time structure alone. By his own admission, he never succeeded in extending the chessboard picture to higher dimensions [89]. He abandoned this program in 60's partially because he felt dissatisfaction with not being able to extend the picture to higher dimensions and partially because Grassmann integral (invented in 60's be Berezin) became just the right tool to describe fermions in line with his path-integral philosophy. In spite of (or perhaps because of) Feynman's abandonment of the checkerboard paradigm, there has been over the years a growing number of works indicating that the checkerboard picture is not merely an interesting mathematical curiosity but it may have more substance than originally thought [142, 225, 240, 241]. For an interested reader I present a technical exposition of the FC in Appendix K.

3.4 Emergent doubly special relativity

As described in the previous subsection, the configuration space for a relativistic particle could be operationally described through polycrystalline or inhomogeneous vacuum. We have also seen that the emergence of special relativity in a QM setting is tightly connected with a particular distribution of grains or regions of inhomogeneity. One might thus ask how robust is this result in regard to a slight perturbation of the mass-smearing distribution. I will now show that the answer is closely related to the concept of *doubly special relativity* or *deformed special relativity* (DSR).

In a nutshell, DSR is a theory which tries to implement coherently a second invariant, besides the speed of light, into the transformations among inertial frames. This new invariant is usually assumed to be an observer-independent length-scale — the Planck length $\ell_p \approx 1.62 \times 10^{-35}$m, or its inverse, i.e., the Planck energy $E_p = c\ell_p^{-1}$. It is thus not surprising that DSR is often employed in connection with various quantum gravity models [102, 103, 274, 292, 293].

To extend my previous reasonings to DSR, I start by considering the modified invariant, or deformed dispersion relation,

$$\frac{\eta^{ab} p_a p_b}{(1 - \ell_p p_0)^2} \;=\; m^2 c^2 \,, \tag{3.25}$$

proposed by Magueijo and Smolin [209, 210, 226]. Here m plays the rôle of the DSR invariant mass. Assuming a metric signature $(+,-,-,-)$, one can solve (3.25) with respect to p_0, which essentially coincides with the

physical Hamiltonian $\bar{H} = cp_0$. The latter is the generator of the temporal translations with respect to the coordinate time t, cf. [209, 210]. My starting Hamiltonian is therefore

$$\bar{H}_{\mp} \;=\; c\,\frac{-m^2c^2\ell_p \mp \sqrt{\mathbf{p}^2(1-m^2c^2\ell_p^2)+m^2c^2}}{1-m^2c^2\ell_p^2}\,, \tag{3.26}$$

and similarly as in the previous subsection only the positive root of $\bar{H}_{\mp}$ is important in superstatistic PI. At this stage I assume that the transformed Hamiltonian entering the PI representation of $\wp(x',t'|x,t)$ is of the form $\bar{H}_+(p,x) = F(H(p,x))$ where F is some Lévy–Khinchine function, which ensures that $H(p,x)$ is non-relativistic Hamiltonan for a free particle. It is not obvious a priori that such F exists. In close analogy with (3.21), it is possible to show [159] that the following superstatistics PI identity holds

$$\begin{aligned}\int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(t)=\mathbf{x}} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\,\exp\left[\int_0^T d\tau\,\left(i\mathbf{p}\cdot\dot{\mathbf{x}}\;+\;\bar{H}_+\right)\right]\\ =\;\int_0^\infty d\tilde{m}\,f_{\frac{1}{2}}\left(\tilde{m},\hat{T}c^2,\hat{T}c^2m^2\right)\\ \times\;\int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(T)=\mathbf{x}}\mathcal{D}\mathbf{x}\,\mathcal{D}\mathbf{p}\,\exp\left[\int_0^T d\tau\,\left(i\mathbf{p}\cdot\dot{\mathbf{x}}-\frac{\mathbf{p}^2}{2\tilde{m}}-E_0\right)\right].\end{aligned} \tag{3.27}$$

Here $E_0 = mc^2/(1+mc\ell_p)$ is the particle's rest energy, see, e.g., [210], $\hat{T} = T\lambda$ and $\lambda = 1/(1-m^2c^2\ell_p^2)$ is the deformation parameter. From the explicit form of the smearing distribution, it is easy to find that $\langle\tilde{m}\rangle = m+1/(\hat{T}c^2)$ and $\mathrm{var}(\tilde{m}) = m/\hat{T}c^2 + 2/\hat{T}^2c^4$.

From the structure of $\langle\tilde{m}\rangle$ we can obtain further useful insights. Similarly as in the special relativistic framework, the fluctuating Newtonian mass $\tilde{m}$ converges for large times T to the DSR invariant mass m. However, now, the rate of convergence is also controlled by the parameter λ. Recalling that $E_p = c/\ell_p$, I can write that $\lambda = 1/(1-E^2/E_p^2)$. So, $\langle\tilde{m}\rangle$ can converge rapidly to the invariant mass m, even at short times T, provided the particle's energy E is close to the Planck energy E_p.

From the identity (3.27) one can deduce the canonical commutation relations (CCR) via the standard PI analysis [92]. In particular, the CCR can be directly related to the degree of roughness (described through Hausdorff dimension D_H or Hurst exponent H) of a sample path in PI [92, 186, 282]. For instance, the usual non-relativistic canonical relation $[\hat{\mathbf{x}}_i,\hat{\mathbf{p}}_j] = i\delta_{ij}$ results from the fact that, for a sample path occurring in non-relativistic PI's,

D_H and H are 3/2 (in time-position plane) and 1/2, respectively. In fact, in non-relativistic quantum mechanics all *local* potentials fall into the same universality class (as for the scaling behavior — $D_H = 3/2$) as the free system [186]. The latter might be viewed as a PI justification of the universal form of non-relativistic CCR's.

It is not hard to show [159] that the PI identity (3.27) implies the CCR

$$[\hat{\mathbf{x}}_i, \hat{\mathbf{p}}_j]_{\rm DSR1} \;=\; i\left(\delta_{ij} \;+\; \frac{\kappa^2 - m^2c^2}{\kappa^2 m^2 c^2}\,\hat{\mathbf{p}}_i\hat{\mathbf{p}}_j\right). \tag{3.28}$$

Here $\kappa = 1/\ell_p$. The CCR (3.28) coincides with the so-called *generalized uncertainty principles* (or GUP) with the GUP *deformation parameter* [278, 279]

$$\beta \;=\; \frac{\kappa^2 - m^2c^2}{m^2c^2}\,. \tag{3.29}$$

We may note in passing that when $mc \to \kappa$, i.e., when m coincides with the Planck mass, i.e. $m = M_p \approx 2.18 \times 10^{-8}$kg then the CCR (3.28) reduces to the non-relativistic one — as expected. This can also be directly seen from (3.27), where for $m \to M_p$ the deformation parameter $\lambda \to \infty$, and the smearing distribution $f_{\frac{1}{2}}\left(\tilde{m}, \hat{T}c^2, \hat{T}c^2m^2\right) \to \delta(m - \tilde{m})$, which yields the usual PI for a Wiener process. On the other hand, when $\kappa \gg mc$ we obtain

$$[\hat{\mathbf{x}}_j, \hat{\mathbf{p}}_i]_{\rm DSR1} \;\approx\; i\left(\delta_{ij} \;+\; \frac{\hat{\mathbf{p}}_i\hat{\mathbf{p}}_j}{m^2c^2}\right). \tag{3.30}$$

which coincides with the special-relativistic commutator [3].

Should I have used instead of (3.25) a different DSR dispersion relation, for example

$$\frac{p_0^2 - \mathbf{p}^2}{1 - (\ell_p p_0)^2} \;=\; m^2c^2\,, \tag{3.31}$$

(which is discussed, e.g., in Ref. [172]), I would have obtained the DSR Hamiltonian

$$\bar{H}_\pm \;=\; \pm\frac{\sqrt{\mathbf{p}^2c^2 + m^2c^4}}{\sqrt{1 + m^2c^2\ell_p^2}}\,, \tag{3.32}$$

from which follows the superstatistics identity holds:

[3]The CCR (3.30) is obtained by lifting Dirac brackets (corresponding to the first class constraint $\Phi \equiv p^2 - m^2$ and the Polyakov gauge condition $\chi \equiv x_\mu p^\mu - \varsigma m^2c^2$, with ς being the world-line parameter) to QM commutators [98, 159]

$$\int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(T)=\mathbf{x}} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, \exp\left\{\int_0^T d\tau\, \left[i\mathbf{p}\cdot\dot{\mathbf{x}} \;-\; \frac{\sqrt{\mathbf{p}^2c^2+m^2c^4}}{\sqrt{1+m^2c^2\ell_p^2}}\right]\right\}$$

$$= \int_0^\infty d\tilde{m}\, f_{\frac{1}{2}}\big(\tilde{m}, Tc^2\zeta^2, Tc^2m^2\big)$$

$$\times \int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(T)=\mathbf{x}} \mathcal{D}\mathbf{x}\,\mathcal{D}\mathbf{p}\, \exp\left\{\int_0^T d\tau\, \left[i\mathbf{p}\cdot\dot{\mathbf{x}} - \frac{\mathbf{p}^2}{2\tilde{m}} - \bar{E}_0\right]\right\}. \quad (3.33)$$

Here $\bar{E}_0 = mc^2/\sqrt{1+m^2c^2\ell_p^2}$ is the particle's rest energy and the deformation parameter now reads $\zeta = 1/\sqrt{1+m^2c^2\ell_p^2}$. In this case one has $\langle\tilde{m}\rangle = m/\zeta + 1/(c^2T\zeta^2)$ and $\mathrm{var}(\tilde{m}) = m/(Tc^2\zeta^3) + 2/(T^2c^4\zeta^4)$. One can also observe that this double-special relativity model does not have the desired property that the fluctuating mass $\tilde{m}$ converges to the DSR invariant mass in the large T. The PI (3.33) implies the CCR [159]

$$[\hat{\mathbf{x}}_j, \hat{\mathbf{p}}_i]_{\mathrm{DSR2}} \;=\; i\bigg(\delta_{ij} \;+\; \frac{\hat{\mathbf{p}}_i\hat{\mathbf{p}}_j}{m^2c^2}\bigg). \quad (3.34)$$

This commutator coincides with the special-relativistic one. This fact is not so surprising since the fractal dimension D_H of a sample path of the DSR2 system coincides with that of special relativity [159].

Let me finally add a few clarifying comments. The qualitative difference in behavior of CCR (3.28) and (3.34) can be traced back to the fact that commutators in (doubly-)special relativity depend on two things: First, the fundamental commutators are essentially the Dirac brackets of the canonical variables that are promoted on the QM level via Dirac's prescription. The explicit definition of the Dirac brackets depends on the choice of a gauge (gauge fixing condition), which for relativistic systems corresponds to choice of a specific physical time. So, the commutation relations are generally gauge fixing dependent in both SR and DSR systems. In the presented approach, the gauge which is implicitly incorporated in the superstatistics PI is the so-called Polyakov gauge [256].

Second, the fundamental commutator $[\hat{\mathbf{x}}_j, \hat{\mathbf{p}}_i]$ depends (through the Jacobi identities) on the whole symplectic structure of the system (and therefore also on the commutator $[\hat{\mathbf{x}}_j, \hat{\mathbf{x}}_i]$, for example). These are not specified by a particular DSR model, but they have to be chosen aside. Of course, one obtains different theories for different choices of $[\hat{\mathbf{x}}_j, \hat{\mathbf{x}}_i]$. In presented

case, I automatically obtain the same fundamental commutators as Ghosh [98] and Mignemi [227] did. In order to satisfy the Jacoby identities one must assume that $\hat{\mathbf{p}}_j$'s commute but $\hat{\mathbf{x}}_j$'s do not commute. For instance, in DSR1 model we should have

$$[\hat{\mathbf{x}}_j, \hat{\mathbf{x}}_i] \;=\; i\,\frac{\kappa^2 - m^2c^2}{\kappa^2 m^2 c^2}\;(\hat{\mathbf{x}}_j\hat{\mathbf{p}}_i - \hat{\mathbf{x}}_i\hat{\mathbf{p}}_j)\;. \tag{3.35}$$

Such a set of commutators closes the so-called *Snyder algebra* [294].

Let me conclude with an important observation. It is possible to show [159] that a slight perturbation in the mass-smearing distribution in aforementioned DSR models would yield again identical DSR systems with slightly perturbed deformation parameter. From this standpoint, the DSR (as well as its low-energy limit — special relativity) is a robust concept, i.e. its algebraic structure continues to hold despite (potentially dynamical) alterations in polycrystalline structure conditions.

3.4.1 *Cosmological implications*

Gravity and Cosmology. — When spacetime is curved, a metric tensor enters in both PI's in (3.21) in a different way, yielding different "counterterms" [20, 179]. I.e., terms that are generated in the PI action action as a perturbation corrections that ensure the finiteness of the PI. For instance, in Bastianelli–van Nieuwenhuizen's time slicing regularization scheme [20] one has (when $\hbar$ is reintroduced)

$$\begin{aligned}
&\frac{\mathbf{p}^2}{2\tilde{m}} \;\mapsto\; \frac{g^{ij}p_ip_j}{2\tilde{m}} \;+\; \frac{\hbar^2}{8\tilde{m}}(R + g^{ij}\Gamma^m_{il}\Gamma^l_{jm})\,,\\
&\sqrt{\mathbf{p}^2 + m^2c^2} \;\mapsto\; \sqrt{g^{ij}p_ip_j \;+\; \frac{\hbar^2}{4}(R \;+\; g^{ij}\Gamma^m_{il}\Gamma^l_{jm}) \;+\; m^2c^2}\\
&+\; \hbar^4\Phi(R, \partial R, \partial^2 R) \;+\; \mathcal{O}(\hbar^6)\,,
\end{aligned} \tag{3.36}$$

where g^{ij}, R, Γ^j_{kl} and $\Phi(\ldots)$ are the (space-like) pull-back metric tensor, the scalar curvature, the Christoffel symbol, and non-vanishing function of R and its first and second derivatives, respectively. This causes the superstatistics identity (3.21) to break down, which one can explicitly check to the leading order in $\hbar$.

The two respective cases have different physical consequences. This implies the breakdown of the superstatistics PI identity. Because the Einstein *equivalence principle* requires that the *local* spacetime structure should be identifiable with the Minkowski spacetime possessing Lorentz symmetry, one might assume the validity of (3.21) at least locally. However, in

different space-time points one has, in general, a different typical length scale of the local inertial frames, depending on the gravitational field. The characteristic size of the local inertial (i.e. Minkowski) frame is of order $1/|K|^{1/4}$ where $K = R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ is the *Kretschmann invariant* and $R_{\alpha\beta\gamma\delta}$ is the Riemann curvature tensor. Relation (3.21) tells us that the special relativistic description breaks down in regions of size smaller than $\lambda_{_C}$.

For curvatures large enough, namely for strong gravitational fields, the size of the local inertial frame can become smaller than $\lambda_{_C}$, that is $1/|K|^{1/4} \lesssim \lambda_{_C}$. In such regions the special relativistic description is no more valid, and according to (3.21) must be replaced by a Newtonian description of the events. For instance, in Schwarzschild geometry[4] one has $K = 12\, r_s^2/r^6$, and the breakdown should be expected at radial distances $r \lesssim (\lambda_{_C}^2 r_s)^{1/3}$ ($r_s = 2Gm/c^2$ is the *Schwarzschild radius*, G is the gravitational constant and m is the source mass) which are — apart from the hypothetical case of micro-black holes (where $\lambda_{_C} \simeq r_s$) — always deeply buried below the Schwarzschild event horizon. In the cosmologically relevant Friedmann–Lemaître–Robertson–Walker (FLRW) geometry[5], one has $K = 12\,(\dot{a}^4 + a^2\ddot{a}^2)/(ac)^4$, and the breakdown happens when $(\dot{a}^4 + a^2\ddot{a}^2) \gtrsim (ac/\lambda_{_C})^4$, where $a(t)$ is the FLRW scale factor of the Universe and $\dot{a} = da/dt$. Applying the Vilenkin–Ford model [325] for inflationary cosmology, where $a(t)$ is given by: $a(t) = A\sqrt{\sinh(Bt)}$ with $B = 2c\sqrt{\Lambda/3}$ (Λ is the cosmological constant), I obtain a temporal bound on the validity of local Lorentz invariance, which, expressed in FLRW time, is

$$t \;\lesssim\; \frac{1}{B}\,\mathrm{arcsinh}\left[\frac{B\lambda_{_C}}{(8c^4 - (B\lambda_{_C})^4)^{1/4}}\right] \;\equiv\; \bar{t}\,. \tag{3.37}$$

By using the presently known value of $\Lambda \simeq 10^{-52}\mathrm{m}^{-2}$ (cf. e.g., Ref. [60]) and the τ-lepton Compton's wavelength $\lambda_{_C}^{\tau} \simeq 6.7 \times 10^{-16}\mathrm{m}$ (yielding the tightest upper bound on t), I obtain $\bar{t} \simeq 4 \times 10^{-24}\mathrm{s}$. Note that, since $B\lambda_{_C} \ll c$, then $\bar{t} \simeq \lambda_{_C}/c = t_{_C}$. Such a violation of the local Lorentz invariance naturally breaks the particle-antiparticle symmetry since there is no unified theory of particles and antiparticles in the non-relativistic physics — formally one has two distinct theories. If the resulting matter-antimatter asymmetry provides a large enough CP asymmetry then this might have essential consequences in the early Universe, e.g., for leptogenesis. In this

[4]The Schwarzschild geometry is described by the Schwarzschild metric, which is the most general spherically symmetric vacuum solution of the Einstein field equations.

[5]The FLRW geometry is characterized by the metric tensor, which describes a homogeneous, isotropic, expanding (or contracting) universe.

respect, $\bar{t}$ is compatible with the *nonthermal* leptogenesis period that typically dates between 10^{-26}–10^{-12}s after the Big Bang.

3.5 Financial markets and econophysics - II

Another important application of the superstatistics paradigm is in *econophysics*. As mentioned in Chapter 2, econophysics is an emerging interdisciplinary field that makes use of concepts and methods of statistical physics in order to analyze complex economical phenomena (be it financial time series, portfolio management or success and failure of companies). The validity and promise of this paradigm has been demonstrated in a number of recent publications and review papers, see e.g., Refs. [45, 179, 222, 307] and citations therein. For the reader's convenience I provide a glossary of relevant financial terms in Supplementary notes.

My take on superstatistics applications in econophysics was presented in three papers [148, 150, 155]. In particular, in Ref. [148] the so-called *stochastic adiabatic theorem* was proved, in Ref. [150] a generalization of the option pricing formula for a Gamma-superstatistics (non-Gaussian) price fluctuations was introduced and discussed, and in Ref. [155] the transitions between distinct superstatistical regimes in the context of high-frequency (minute-tick) share-price returns of seven selected companies was observed and analyzed.

Let me start first with the stochastic adiabatic theorem.

3.5.1 *Stochastic "adiabatic" theorem:*

As seen in Section 3.2 one can compute the conditional probabilities for Markovian processes by employing PI calculus. For non-quadratic Hamiltonians such computations are typically quite difficult and hence one might prefer to study such conditional probabilities directly in terms of their time-evolution equations. Such equations are called (forward) *Kramers–Moyal equations* and are of the form [269]

$$\partial_{t'} P(x', t'|x, t) \;=\; \mathbb{L} P(x', t'|x, t)\,, \tag{3.38}$$

where the Kramers–Moyal operator $\mathbb{L}$ has the expansion

$$\mathbb{L}(\partial_{t'}, x') \;=\; \sum_{n=1}^{\infty} (-\partial_{x'})^n D^{(n)}(x', t')\,. \tag{3.39}$$

Here the coefficients $D^{(n)}(x',t')$ are equal to the moments of the short-time conditional probability $P(x',t'|x,t)$, namely

$$\begin{aligned} D^{(n)}(x,t) &= \frac{1}{n!}\,\lim_{\tau\to 0}\,\frac{1}{\tau}\int_{-\infty}^{\infty} dz\,(z-x)^n\,P(z,t+\tau|x,t) \\ &= \frac{1}{n!}\,\lim_{\tau\to 0}\,\frac{1}{\tau}\int_{-\infty}^{\infty} dz\,(z-x)^n\,\langle z|e^{-\tau\hat{H}}|y\rangle\,, \end{aligned} \tag{3.40}$$

cf. e.g., [148, 269]. According to *Pawula's theorem* [249] (see also Appendix J) there is either infinite number of non-zero coefficients $D^{(n)}(x,t)$ or, if there is a finite number of them, they can be non-zero only up to $n=2$. In addition, any artificial truncation of of expansion (3.39) at $n>2$ would produce non-positive conditional probabilities. This is the basic reason why phenomenological models for $\mathbb{L}$ go typically only up to $n=2$, i.e. to Fokker–Planck operator $\mathbb{L}^{\rm FP}$ and ensuing (forward) *Fokker–Planck equation*.

Because Eq. (3.38) involves derivatives with respect to final coordinates (x',t'), it is also called *forward* Kramers–Moyal equation. To solve this equation one has to specify boundary conditions at the initial time t. Since $\hat{H}$ might contain diffusive terms, e.g. terms proportional to momenta, backward time evolution is generally not described by the same operator $\mathbb{L}$ as the forward evolution. Though the backward Kramers–Moyal equations are easy to construct (see e.g. [269]), they are less relevant in practical applications, and hence I shall deal here only with their forward counterparts.

Let me first start with the following theorem.

Theorem 3.1 ("Adiabatic" theorem). *If both* $P_\zeta(x',t'|x,t)$ *and*

$$P(x',t'|x,t) \;=\; \int d\zeta\; P_\zeta(x',t'|x,t)\,\omega(\zeta,T) \quad \textit{with} \quad T=t'-t\,,$$

are Markovian PDF's then their dynamics can be described via two (forward) Fokker–Planck equations:

$$\begin{aligned} \partial_T\,\omega(\zeta,T) &= \mathbb{L}^{\rm FP}_\omega\,\omega(\zeta,T)\,, \\ \partial_{t'}P_\zeta(x',t'|x,t) &= \mathbb{L}^{\rm FP}_\zeta P_\zeta(x',t'|x,t)\,, \end{aligned} \tag{3.41}$$

with

$$\begin{aligned} \mathbb{L}^{\rm FP}_\omega &= -\partial_\zeta\,K^{(1)}(\zeta,T) \;+\; \partial^2_\zeta\,K^{(2)}(\zeta,T)\,, \\ \mathbb{L}^{\rm FP}_\zeta &= -\partial_{x'}\,D^{(1)}_\zeta(x',t') \;+\; \partial^2_{x'}\,D^{(2)}_\zeta(x',t')\,. \end{aligned} \tag{3.42}$$

Here $K^{(1)}, D_\zeta^{(1)}$ are drift coefficients and $K^{(2)}, D_\zeta^{(2)}$ are diffusion coefficients. In addition, the previous Fokker–Planck equations are on the level of sample trajectories equivalent to two coupled Itô stochastic equations

$$\begin{aligned} dx' &= D_\zeta^{(1)}(x',t')\,dt' + \sqrt{2D_\zeta^{(2)}(x',t')}\,dW_1\,, \\ d\zeta &= K^{(1)}(\zeta,T)\,dT + \sqrt{2K^{(2)}(\zeta,T)}\,dW_2\,. \end{aligned} \tag{3.43}$$

Here $W_1(t')$ and $W_2(T)$ are Wiener processes.

Proof of the theorem is rather lengthy and I would refer the interested reader to the original paper [148]. Let me however, add few comments that would clarify the meaning of Theorem 3.1.

One way to study implications of the superstatistics PI (3.11) is to directly compute such PIs. This is, however, often very hard and aside from simple cases (cf. e.g. Ref. [113]), such computations is accessible only through perturbative methods. Often more easier is to solve associated Kramers–Moyal or Fokker–Planck equations. Since both $P_\zeta(x',t'|x,t)$ and $P(x',t'|x,t)$ satisfy the Chapman–Kolmogorov equation, it would be natural to study Kramers–Moyal or Fokker–Planck equations associated with those transition probabilitis. This might still be quite difficult task due to (typically) non-linear nature of the emergent Hamiltonian $\mathcal{H} = F(H)$. The adiabatic theorem allows to circumvent this by solving equivalent (but often simpler) problem in terms of two coupled Fokker–Planck equations (3.41). In the latter case, the drift and diffusion coefficients $K^{(1)}$ and $K^{(2)}$, respectively, can be easily computed from first two moments of the distribution $\omega(\zeta,T)$, which are often even tabulated.

Notice, further, that there are two very different characteristic time scales involved in Eqs. (3.41): t, the "internal" time scale, representing typical time scale over which $P_\zeta(x',t'|x,t)$ changes (e.g. financial data tick scale, mean collision time or time-slicing scale), and T, the "external" time, over which the "external" parameter ζ changes. Processes where $T \gg t$, as in our case, are called *adiabatic processes* — hence the adjective *adiabatic* in name of the theorem. Basic strategy for analysing an adiabatic process is first to solve the problem with the external parameters held constant, and only at the end of the calculation allow them to vary (slowly) with time. This makes the theorem useful for practical applications.

3.5.2 *Double-stochastic process in financial markets*

The previous adiabatic theorem has a direct implications in financial markets. In particular, it implies that *Gamma superstatistics* PI is closely related to the Heston stochastic volatility model. To see this more explicitly let me start with the Gamma superstatistics PI, i.e. I assume that the distribution $w(\zeta, T)$ is a Gamma distribution with parameters $b > 0$ and $c > 0$, such that

$$w(\zeta, t) \;=\; \frac{(b)^{ct}\zeta^{ct-1}}{\Gamma(ct)}\, e^{-b\zeta}\,, \qquad \zeta \in \mathbb{R}^+\,, \tag{3.44}$$

or in terms of the smearing distribution $\omega(\zeta, t)$

$$\omega(\zeta, t) \;=\; \frac{(bt)^{ct}\zeta^{ct-1}}{\Gamma(ct)}\, e^{-bt\zeta}\,, \qquad \zeta \in \mathbb{R}^+\,. \tag{3.45}$$

Since a Gamma distribution is infinitely divisible [84], it has the Lévy–Khintchine representation of the moment-generating function cf. Eq. (3.15). It can be directly checked that in this case the Lévy measure $\mu(du) = ce^{-bu}u^{-1}\mathbf{1}_{(0,+\infty)}(u)du$ and $\alpha = 0$. With this we get

$$F(p_\zeta) \;=\; \bar{\zeta}b\,\left[\log\Big(\frac{p_\zeta}{b} \;+\; 1\Big)\right], \tag{3.46}$$

where $\bar{\zeta} = c/b$ is the mean of $\omega(\zeta, t)$. In particular, for $H = \mathbf{p}^2/2$ we have

$$\mathcal{H}(p) \;=\; \bar{\zeta}b\,\left[\log\left(\frac{\mathbf{p}^2}{2b} \;+\; 1\right)\right], \tag{3.47}$$

which leads to the superstatistics PI identity

$$\begin{aligned}
\wp(x', t'|x, t) &= \int_{\mathbf{x}(t)\,=\,\mathbf{x}}^{\mathbf{x}(t')=\mathbf{x}'} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, e^{\int_t^{t'} d\tau\left[i\mathbf{p}\dot{\mathbf{x}} \;-\; \bar{\zeta}b\log\left(\mathbf{p}^2/2b+1\right)\right]} \\
&= \int_0^\infty d\zeta\, \omega(\zeta, T) \int_{\mathbf{x}(0)\,=\,\mathbf{x}}^{\mathbf{x}(T)\,=\,\mathbf{x}'} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, e^{\int_t^{t'} d\tau(i\mathbf{p}\dot{\mathbf{x}} - \zeta\mathbf{p}^2/2)} \\
&\equiv \int_0^\infty d\zeta\, \omega(\zeta, T)\, \wp_\zeta(x', t'|x, t)\,,
\end{aligned} \tag{3.48}$$

with $T = t' - t$. Since the RHS of (3.48) involves only PI with quadratic action, I can integrate it out to obtain

$$\int_0^\infty d\zeta\, \frac{(bT)^{cT}\zeta^{cT-1}}{\Gamma(cT)} \sqrt{\frac{1}{2\pi T\zeta}}\, e^{-bT\zeta} e^{-\mathbf{r}^2/2\zeta T}\,, \tag{3.49}$$

where $\mathbf{r} = \mathbf{x}' - \mathbf{x}$. After the integration I find

$$K_{1/2-cT}(2\sqrt{2b}\,|\mathbf{r}|)\,\frac{T^{-3/2}}{\sqrt{\pi}\,\Gamma(cT)}\left(\frac{2\sqrt{2b}\,T}{|\mathbf{r}|}\right)^{1/2-cT}. \tag{3.50}$$

It is quite interesting to note that in Fourier space, this coincides with a Tsallis distribution [24, 25, 150, 333].

Connection with theory of financial markets arises when I consider the free-particle Hamiltonian with drift: $p^2/2 + ip(r/\zeta - 1/2)$, see e.g., Refs. [150, 179]. Here r is the *drift constant* and ζ *volatility* (see Supplementary notes H.1). In order to find forward Fokker–Planck equations for $\wp_\zeta(x', t'|x, t)$ and $\omega(\zeta, T)$, I need to compute corresponding *drift* and *diffusion coefficients* $D_\zeta^{(1)}$ and $D_\zeta^{(2)}$ for $\wp_\zeta(x', t'|x, t)$ and similarly $K^{(1)}$ and $K^{(2)}$ for $\omega(\zeta, T)$. By employing the prescription of Eq. (3.40), the drift coefficients read

$$D_\zeta^{(1)} \;=\; \left(r - \frac{\zeta}{2}\right), \quad D_\zeta^{(2)}(x', t') \;=\; \frac{\zeta}{2}\,. \tag{3.51}$$

Similarly, the diffusion coefficients $K^{(n)}$ are

$$\begin{aligned} K^{(1)}(\zeta, T) \;&=\; \frac{1}{T}\left(\frac{c}{b} - \zeta\right) \;=\; \frac{1}{T}\left(\bar{\zeta} - \zeta\right), \\ K^{(n)}(\zeta, T) \;&=\; \frac{1}{T^n}\frac{c}{nb^n}\,, \qquad n \geq 2\,. \end{aligned} \tag{3.52}$$

With the help of adiabatic theorem I can write down the underlying Itō processes for corresponding sample trajectories, namely

$$\begin{aligned} dx' \;&=\; \left(r - \frac{\zeta}{2}\right) dt' \;+\; \sqrt{\zeta}\, dW_1 \\ d\zeta \;&=\; \frac{1}{T}\left(\bar{\zeta} - \zeta\right) dT \;+\; \frac{1}{T}\sqrt{\frac{\bar{\zeta}}{b}}\, dW_2\,. \end{aligned} \tag{3.53}$$

I can now view x' as a logarithm of a stock price S, and ζ and r as the associated variance and drift. If, in addition, I replace for large T the quantity $\sqrt{\bar{\zeta}}$ with $\sqrt{\zeta}$, the systems (3.53) reduces to

$$\begin{aligned} dS \;&=\; rS\, dt' \;+\; \sqrt{\zeta} S\, dW_1\,, \\ d\zeta \;&=\; \gamma\left(\bar{\zeta} - \zeta\right) dT + \varepsilon\sqrt{\zeta}\; dW_2\,, \end{aligned} \tag{3.54}$$

where $\gamma = 1/T$ and $\varepsilon = 1/(\sqrt{b}T)$. The system of equations (3.54) constitute Heston's stochastic volatility model [126] which is often used in financial

markets to model bond and currency options (for further reference on this model see, e.g., Ref. [126, 179] and citations therein). In this framework, the parameters $\bar{\zeta}$, γ and ε are interpreted as the long-time volatility average, the drift of the variance and the volatility of the variance, respectively.

Let us finally observe that in the large T limit $\omega(\zeta, T) = \delta(\zeta - \bar{\zeta})$ and so volatility is constant. In this case the solution of the Focker–Planck equation (3.41) with the initial condition $\wp_{\bar{\zeta}}(x', t|x, t) = \delta(x' - x)$ can be easily found as follows. I define first the short-time transition probability $\wp_{\bar{\zeta}}(x', t + \Delta t|x, t)$ where time increment Δt is small. With this I can write

$$\begin{aligned} \wp_{\bar{\zeta}}(x', t + \Delta t|x, t) &= \exp(\tau \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}})\, \wp_{\bar{\zeta}}(x', t|x, t) \\ &= \exp(\tau \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}})\, \delta(x' - x)\,. \end{aligned} \tag{3.55}$$

Now I expand $\exp(\tau \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}})$ in Δt, which gives

$$\begin{aligned} \wp_{\bar{\zeta}}(x', t + \Delta t|x, t) &= \left[1 + \Delta t \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}} + \mathcal{O}(\Delta t^2)\right] \delta(x' - x) \\ &= \left[1 - \Delta t \partial_x\, D_{\bar{\zeta}}^{(1)}(x, t) \;+\; \Delta t \partial_x^2\, D_{\bar{\zeta}}^{(2)}(x, t) + \mathcal{O}(\Delta t^2)\right] \delta(x' - x). \end{aligned} \tag{3.56}$$

On the second line I have exchanged x for x' inside of $[\ldots]$. By using the Fourier transform representation of the δ-function, I can further rewrite the RHS of (3.56) as

$$\begin{aligned} &\frac{1}{2\pi} \int_{-\infty}^{\infty} dp\, \left[1 - ip\Delta t D_{\bar{\zeta}}^{(1)}(x, t) \;-\; p^2 \Delta t D_{\bar{\zeta}}^{(2)}(x, t) + \mathcal{O}(\Delta t^2)\right] e^{ip(x' - x)} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp\, \exp\left[ip(x' - x) - ip\Delta t D_{\bar{\zeta}}^{(1)}(x, t) \;-\; p^2 \Delta t D_{\bar{\zeta}}^{(2)}(x, t)\right], \end{aligned} \tag{3.57}$$

where on the first line integration by parts was used and on the second line the contributions of the order $\mathcal{O}(\Delta t^2)$ were neglected. Integration of the Gaussian integral yields the short-time transition probability

$$\begin{aligned} &\wp_{\bar{\zeta}}(x', t + \Delta t|x, t) \\ &\quad = \frac{1}{2\sqrt{\pi \Delta t D_{\bar{\zeta}}^{(2)}(x, t)}} \exp\left\{-\frac{[(x' - x) - \Delta t D_{\bar{\zeta}}^{(1)}(x, t)]^2}{4\Delta t D_{\bar{\zeta}}^{(2)}(x, t)}\right\} \\ &\quad = \frac{1}{\sqrt{2\pi \Delta t \bar{\zeta}}} \exp\left\{-\frac{[(x' - x) - \Delta t(r - \bar{\zeta}/2)]^2}{2\Delta t \bar{\zeta}}\right\}. \end{aligned} \tag{3.58}$$

The short-time transition probability can be now used to find a final-time solution of the Fokker–Planck equation. To this end I divide time difference

T into $N+1$ small intervals of length $\Delta t = T/(N+1)$, defining $t_n = t + n\tau$ (so that $t = t_0$ and $t' = t_{N+1}$), and by repeatedly applying the Chapman–Kolmogorov equation (3.10), I get

$$\begin{aligned}\wp_{\bar{\zeta}}(x',t'|x,t) &= \int_{\mathbb{R}} dx_{_N} \cdots \int_{\mathbb{R}} dx_1\, \wp_{\bar{\zeta}}(x',t'|x_{_N},t_{_N}) \cdots \wp_{\bar{\zeta}}(x_1,t_1|x,t) \\ &= \left[\sqrt{\frac{1}{2\pi\Delta t\bar{\zeta}}} \prod_{k=1}^{N} \int_{\mathbb{R}} \left(dx_k \sqrt{\frac{1}{2\pi\Delta t\bar{\zeta}}} \right) \right] \\ &\times \exp\left\{ -\sum_{n=0}^{N} \frac{[(x_{n+1}-x_n) - \Delta t(r-\bar{\zeta}/2)]^2}{2\Delta t\bar{\zeta}} \right\}, \end{aligned} \tag{3.59}$$

which in the large N limit tends to the PI (cf. Appendix I)

$$\begin{aligned}&\int_{x(t)=x}^{x(t')=x'} \mathcal{D}x\ \exp\left\{ -\int_t^{t'} dt\, \frac{[\dot{x} - (r-\bar{\zeta}/2)]^2}{2\bar{\zeta}} \right\} \\ &= \int_{x(t)=x}^{x(t')=x'} \mathcal{D}x\mathcal{D}p \exp\left\{ -\int_t^{t'} dt\, \left[ip\dot{x} - \frac{\bar{\zeta}p^2}{2} - ip\left(r - \frac{\bar{\zeta}}{2}\right) \right] \right\} \\ &= \int_{x(t)=x}^{x(t')=x'} \mathcal{D}x\mathcal{D}p \exp\left\{ -\int_t^{t'} dt\, \left(ip\dot{x} - \bar{\zeta}H \right) \right\}, \end{aligned} \tag{3.60}$$

with $H = p^2/2 + ip(r/\zeta - 1/2)$. This PI representation precisely coincides with the prescription (3.48) where instead of a free-particle Hamiltonian I use the Hamiltonian with drift H. Since this PI has a well known solution [113], the full transition probability can be written in a closed form

$$\wp_{\bar{\zeta}}(x',t'|x,t) = \frac{1}{\sqrt{2\pi\bar{\zeta}T}} \exp\left\{ -\frac{\left[(x'-x) - \left(r - \frac{\bar{\zeta}}{2}\right) T \right]^2}{2\bar{\zeta}T} \right\}. \tag{3.61}$$

The transition probability distribution (3.61) is recognized as the risk-free member of the family of Gaussian martingale distributions for the stock price $S(t) = e^{x(t)}$, cf. e.g. Ref. [179].

moments of time is equal to λ_C. Feynman hoped that with FC he could explain particle's spin as a result of the space-time structure alone. By his own admission, he never succeeded in extending the chessboard picture to higher dimensions [89]. He abandoned this program in 60's partially because he felt dissatisfaction with not being able to extend the picture to higher dimensions and partially because Grassmann integral (invented in 60's be Berezin) became just the right tool to describe fermions in line with his path-integral philosophy. In spite of (or perhaps because of) Feynman's abandonment of the checkerboard paradigm, there has been over the years a growing number of works indicating that the checkerboard picture is not merely an interesting mathematical curiosity but it may have more substance than originally thought [142, 225, 240, 241]. For an interested reader I present a technical exposition of the FC in Appendix K.

3.4 Emergent doubly special relativity

As described in the previous subsection, the configuration space for a relativistic particle could be operationally described through polycrystalline or inhomogeneous vacuum. We have also seen that the emergence of special relativity in a QM setting is tightly connected with a particular distribution of grains or regions of inhomogeneity. One might thus ask how robust is this result in regard to a slight perturbation of the mass-smearing distribution. I will now show that the answer is closely related to the concept of *doubly special relativity* or *deformed special relativity* (DSR).

In a nutshell, DSR is a theory which tries to implement coherently a second invariant, besides the speed of light, into the transformations among inertial frames. This new invariant is usually assumed to be an observer-independent length-scale — the Planck length $\ell_p \approx 1.62 \times 10^{-35}$m, or its inverse, i.e., the Planck energy $E_p = c\ell_p^{-1}$. It is thus not surprising that DSR is often employed in connection with various quantum gravity models [102, 103, 274, 292, 293].

To extend my previous reasonings to DSR, I start by considering the modified invariant, or deformed dispersion relation,

$$\frac{\eta^{ab} p_a p_b}{(1 - \ell_p p_0)^2} \;=\; m^2 c^2 \,, \tag{3.25}$$

proposed by Magueijo and Smolin [209, 210, 226]. Here m plays the rôle of the DSR invariant mass. Assuming a metric signature $(+,-,-,-)$, one can solve (3.25) with respect to p_0, which essentially coincides with the

physical Hamiltonian $\bar{H} = cp_0$. The latter is the generator of the temporal translations with respect to the coordinate time t, cf. [209, 210]. My starting Hamiltonian is therefore

$$\bar{H}_{\mp} \;=\; c\,\frac{-m^2c^2\ell_p \mp \sqrt{\mathbf{p}^2(1-m^2c^2\ell_p^2)+m^2c^2}}{1-m^2c^2\ell_p^2}\,, \tag{3.26}$$

and similarly as in the previous subsection only the positive root of $\bar{H}_{\mp}$ is important in superstatistic PI. At this stage I assume that the transformed Hamiltonian entering the PI representation of $\wp(x',t'|x,t)$ is of the form $\bar{H}_+(p,x) = F(H(p,x))$ where F is some Lévy–Khinchine function, which ensures that $H(p,x)$ is non-relativistic Hamiltonan for a free particle. It is not obvious a priori that such F exists. In close analogy with (3.21), it is possible to show [159] that the following superstatistics PI identity holds

$$\begin{aligned}\int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(t)=\mathbf{x}} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\,\exp\left[\int_0^T d\tau\,\left(i\mathbf{p}\cdot\dot{\mathbf{x}}\;+\;\bar{H}_+\right)\right]\\ =\;\int_0^\infty d\tilde{m}\,f_{\frac{1}{2}}\left(\tilde{m},\hat{T}c^2,\hat{T}c^2m^2\right)\\ \times\;\int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(T)=\mathbf{x}}\mathcal{D}\mathbf{x}\,\mathcal{D}\mathbf{p}\,\exp\left[\int_0^T d\tau\,\left(i\mathbf{p}\cdot\dot{\mathbf{x}}-\frac{\mathbf{p}^2}{2\tilde{m}}-E_0\right)\right].\end{aligned} \tag{3.27}$$

Here $E_0 = mc^2/(1+mc\ell_p)$ is the particle's rest energy, see, e.g., [210], $\hat{T} = T\lambda$ and $\lambda = 1/(1-m^2c^2\ell_p^2)$ is the deformation parameter. From the explicit form of the smearing distribution, it is easy to find that $\langle\tilde{m}\rangle = m+1/(\hat{T}c^2)$ and $\mathrm{var}(\tilde{m}) = m/\hat{T}c^2 + 2/\hat{T}^2c^4$.

From the structure of $\langle\tilde{m}\rangle$ we can obtain further useful insights. Similarly as in the special relativistic framework, the fluctuating Newtonian mass $\tilde{m}$ converges for large times T to the DSR invariant mass m. However, now, the rate of convergence is also controlled by the parameter λ. Recalling that $E_p = c/\ell_p$, I can write that $\lambda = 1/(1-E^2/E_p^2)$. So, $\langle\tilde{m}\rangle$ can converge rapidly to the invariant mass m, even at short times T, provided the particle's energy E is close to the Planck energy E_p.

From the identity (3.27) one can deduce the canonical commutation relations (CCR) via the standard PI analysis [92]. In particular, the CCR can be directly related to the degree of roughness (described through Hausdorff dimension D_H or Hurst exponent H) of a sample path in PI [92, 186, 282]. For instance, the usual non-relativistic canonical relation $[\hat{\mathbf{x}}_i,\hat{\mathbf{p}}_j] = i\delta_{ij}$ results from the fact that, for a sample path occurring in non-relativistic PI's,

D_H and H are 3/2 (in time-position plane) and 1/2, respectively. In fact, in non-relativistic quantum mechanics all *local* potentials fall into the same universality class (as for the scaling behavior — $D_H = 3/2$) as the free system [186]. The latter might be viewed as a PI justification of the universal form of non-relativistic CCR's.

It is not hard to show [159] that the PI identity (3.27) implies the CCR

$$[\hat{\mathbf{x}}_i, \hat{\mathbf{p}}_j]_{\rm DSR1} \;=\; i\left(\delta_{ij} \;+\; \frac{\kappa^2 - m^2c^2}{\kappa^2 m^2 c^2}\,\hat{\mathbf{p}}_i\hat{\mathbf{p}}_j\right). \tag{3.28}$$

Here $\kappa = 1/\ell_p$. The CCR (3.28) coincides with the so-called *generalized uncertainty principles* (or GUP) with the GUP *deformation parameter* [278, 279]

$$\beta \;=\; \frac{\kappa^2 - m^2c^2}{m^2c^2}\,. \tag{3.29}$$

We may note in passing that when $mc \to \kappa$, i.e., when m coincides with the Planck mass, i.e. $m = M_p \approx 2.18 \times 10^{-8}$kg then the CCR (3.28) reduces to the non-relativistic one — as expected. This can also be directly seen from (3.27), where for $m \to M_p$ the deformation parameter $\lambda \to \infty$, and the smearing distribution $f_{\frac{1}{2}}\left(\tilde{m}, \hat{T}c^2, \hat{T}c^2m^2\right) \to \delta(m - \tilde{m})$, which yields the usual PI for a Wiener process. On the other hand, when $\kappa \gg mc$ we obtain

$$[\hat{\mathbf{x}}_j, \hat{\mathbf{p}}_i]_{\rm DSR1} \;\approx\; i\left(\delta_{ij} \;+\; \frac{\hat{\mathbf{p}}_i\hat{\mathbf{p}}_j}{m^2c^2}\right). \tag{3.30}$$

which coincides with the special-relativistic commutator [3].

Should I have used instead of (3.25) a different DSR dispersion relation, for example

$$\frac{p_0^2 - \mathbf{p}^2}{1 - (\ell_p p_0)^2} \;=\; m^2c^2\,, \tag{3.31}$$

(which is discussed, e.g., in Ref. [172]), I would have obtained the DSR Hamiltonian

$$\bar{H}_\pm \;=\; \pm\frac{\sqrt{\mathbf{p}^2c^2 + m^2c^4}}{\sqrt{1 + m^2c^2\ell_p^2}}\,, \tag{3.32}$$

from which follows the superstatistics identity holds:

[3]The CCR (3.30) is obtained by lifting Dirac brackets (corresponding to the first class constraint $\Phi \equiv p^2 - m^2$ and the Polyakov gauge condition $\chi \equiv x_\mu p^\mu - \varsigma m^2c^2$, with ς being the world-line parameter) to QM commutators [98, 159]

$$\int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(T)=\mathbf{x}} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, \exp\left\{\int_0^T d\tau\, \left[i\mathbf{p}\cdot\dot{\mathbf{x}} \;-\; \frac{\sqrt{\mathbf{p}^2c^2+m^2c^4}}{\sqrt{1+m^2c^2\ell_p^2}}\right]\right\}$$

$$= \int_0^\infty d\tilde{m}\, f_{\frac{1}{2}}\big(\tilde{m}, Tc^2\zeta^2, Tc^2m^2\big)$$

$$\times \int_{\mathbf{x}(0)=\mathbf{x}'}^{\mathbf{x}(T)=\mathbf{x}} \mathcal{D}\mathbf{x}\,\mathcal{D}\mathbf{p}\, \exp\left\{\int_0^T d\tau\, \left[i\mathbf{p}\cdot\dot{\mathbf{x}} - \frac{\mathbf{p}^2}{2\tilde{m}} - \bar{E}_0\right]\right\}. \quad (3.33)$$

Here $\bar{E}_0 = mc^2/\sqrt{1+m^2c^2\ell_p^2}$ is the particle's rest energy and the deformation parameter now reads $\zeta = 1/\sqrt{1+m^2c^2\ell_p^2}$. In this case one has $\langle\tilde{m}\rangle = m/\zeta + 1/(c^2T\zeta^2)$ and $\mathrm{var}(\tilde{m}) = m/(Tc^2\zeta^3) + 2/(T^2c^4\zeta^4)$. One can also observe that this double-special relativity model does not have the desired property that the fluctuating mass $\tilde{m}$ converges to the DSR invariant mass in the large T. The PI (3.33) implies the CCR [159]

$$[\hat{\mathbf{x}}_j, \hat{\mathbf{p}}_i]_{\mathrm{DSR2}} \;=\; i\bigg(\delta_{ij} \;+\; \frac{\hat{\mathbf{p}}_i\hat{\mathbf{p}}_j}{m^2c^2}\bigg). \quad (3.34)$$

This commutator coincides with the special-relativistic one. This fact is not so surprising since the fractal dimension D_H of a sample path of the DSR2 system coincides with that of special relativity [159].

Let me finally add a few clarifying comments. The qualitative difference in behavior of CCR (3.28) and (3.34) can be traced back to the fact that commutators in (doubly-)special relativity depend on two things: First, the fundamental commutators are essentially the Dirac brackets of the canonical variables that are promoted on the QM level via Dirac's prescription. The explicit definition of the Dirac brackets depends on the choice of a gauge (gauge fixing condition), which for relativistic systems corresponds to choice of a specific physical time. So, the commutation relations are generally gauge fixing dependent in both SR and DSR systems. In the presented approach, the gauge which is implicitly incorporated in the superstatistics PI is the so-called Polyakov gauge [256].

Second, the fundamental commutator $[\hat{\mathbf{x}}_j, \hat{\mathbf{p}}_i]$ depends (through the Jacobi identities) on the whole symplectic structure of the system (and therefore also on the commutator $[\hat{\mathbf{x}}_j, \hat{\mathbf{x}}_i]$, for example). These are not specified by a particular DSR model, but they have to be chosen aside. Of course, one obtains different theories for different choices of $[\hat{\mathbf{x}}_j, \hat{\mathbf{x}}_i]$. In presented

case, I automatically obtain the same fundamental commutators as Ghosh [98] and Mignemi [227] did. In order to satisfy the Jacoby identities one must assume that $\hat{\mathbf{p}}_j$'s commute but $\hat{\mathbf{x}}_j$'s do not commute. For instance, in DSR1 model we should have

$$[\hat{\mathbf{x}}_j, \hat{\mathbf{x}}_i] \;=\; i\,\frac{\kappa^2 - m^2c^2}{\kappa^2 m^2 c^2}\;(\hat{\mathbf{x}}_j\hat{\mathbf{p}}_i - \hat{\mathbf{x}}_i\hat{\mathbf{p}}_j)\;. \tag{3.35}$$

Such a set of commutators closes the so-called *Snyder algebra* [294].

Let me conclude with an important observation. It is possible to show [159] that a slight perturbation in the mass-smearing distribution in aforementioned DSR models would yield again identical DSR systems with slightly perturbed deformation parameter. From this standpoint, the DSR (as well as its low-energy limit — special relativity) is a robust concept, i.e. its algebraic structure continues to hold despite (potentially dynamical) alterations in polycrystalline structure conditions.

3.4.1 *Cosmological implications*

Gravity and Cosmology. — When spacetime is curved, a metric tensor enters in both PI's in (3.21) in a different way, yielding different "counterterms" [20, 179]. I.e., terms that are generated in the PI action action as a perturbation corrections that ensure the finiteness of the PI. For instance, in Bastianelli–van Nieuwenhuizen's time slicing regularization scheme [20] one has (when $\hbar$ is reintroduced)

$$\begin{aligned}
&\frac{\mathbf{p}^2}{2\tilde{m}} \;\mapsto\; \frac{g^{ij}p_ip_j}{2\tilde{m}} \;+\; \frac{\hbar^2}{8\tilde{m}}(R + g^{ij}\Gamma^m_{il}\Gamma^l_{jm})\,,\\
&\sqrt{\mathbf{p}^2 + m^2c^2} \;\mapsto\; \sqrt{g^{ij}p_ip_j \;+\; \frac{\hbar^2}{4}(R \;+\; g^{ij}\Gamma^m_{il}\Gamma^l_{jm}) \;+\; m^2c^2}\\
&+\; \hbar^4\Phi(R, \partial R, \partial^2 R) \;+\; \mathcal{O}(\hbar^6)\,,
\end{aligned} \tag{3.36}$$

where g^{ij}, R, Γ^j_{kl} and $\Phi(\ldots)$ are the (space-like) pull-back metric tensor, the scalar curvature, the Christoffel symbol, and non-vanishing function of R and its first and second derivatives, respectively. This causes the superstatistics identity (3.21) to break down, which one can explicitly check to the leading order in $\hbar$.

The two respective cases have different physical consequences. This implies the breakdown of the superstatistics PI identity. Because the Einstein *equivalence principle* requires that the *local* spacetime structure should be identifiable with the Minkowski spacetime possessing Lorentz symmetry, one might assume the validity of (3.21) at least locally. However, in

different space-time points one has, in general, a different typical length scale of the local inertial frames, depending on the gravitational field. The characteristic size of the local inertial (i.e. Minkowski) frame is of order $1/|K|^{1/4}$ where $K = R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ is the *Kretschmann invariant* and $R_{\alpha\beta\gamma\delta}$ is the Riemann curvature tensor. Relation (3.21) tells us that the special relativistic description breaks down in regions of size smaller than $\lambda_{_C}$.

For curvatures large enough, namely for strong gravitational fields, the size of the local inertial frame can become smaller than $\lambda_{_C}$, that is $1/|K|^{1/4} \lesssim \lambda_{_C}$. In such regions the special relativistic description is no more valid, and according to (3.21) must be replaced by a Newtonian description of the events. For instance, in Schwarzschild geometry[4] one has $K = 12\, r_s^2/r^6$, and the breakdown should be expected at radial distances $r \lesssim (\lambda_{_C}^2 r_s)^{1/3}$ ($r_s = 2Gm/c^2$ is the *Schwarzschild radius*, G is the gravitational constant and m is the source mass) which are — apart from the hypothetical case of micro-black holes (where $\lambda_{_C} \simeq r_s$) — always deeply buried below the Schwarzschild event horizon. In the cosmologically relevant Friedmann–Lemaître–Robertson–Walker (FLRW) geometry[5], one has $K = 12\,(\dot{a}^4 + a^2\ddot{a}^2)/(ac)^4$, and the breakdown happens when $(\dot{a}^4 + a^2\ddot{a}^2) \gtrsim (ac/\lambda_{_C})^4$, where $a(t)$ is the FLRW scale factor of the Universe and $\dot{a} = da/dt$. Applying the Vilenkin–Ford model [325] for inflationary cosmology, where $a(t)$ is given by: $a(t) = A\sqrt{\sinh(Bt)}$ with $B = 2c\sqrt{\Lambda/3}$ (Λ is the cosmological constant), I obtain a temporal bound on the validity of local Lorentz invariance, which, expressed in FLRW time, is

$$t \;\lesssim\; \frac{1}{B}\,\mathrm{arcsinh}\left[\frac{B\lambda_{_C}}{(8c^4 - (B\lambda_{_C})^4)^{1/4}}\right] \;\equiv\; \bar{t}\,. \tag{3.37}$$

By using the presently known value of $\Lambda \simeq 10^{-52}\mathrm{m}^{-2}$ (cf. e.g., Ref. [60]) and the τ-lepton Compton's wavelength $\lambda_{_C}^{\tau} \simeq 6.7 \times 10^{-16}\mathrm{m}$ (yielding the tightest upper bound on t), I obtain $\bar{t} \simeq 4 \times 10^{-24}\mathrm{s}$. Note that, since $B\lambda_{_C} \ll c$, then $\bar{t} \simeq \lambda_{_C}/c = t_{_C}$. Such a violation of the local Lorentz invariance naturally breaks the particle-antiparticle symmetry since there is no unified theory of particles and antiparticles in the non-relativistic physics — formally one has two distinct theories. If the resulting matter-antimatter asymmetry provides a large enough CP asymmetry then this might have essential consequences in the early Universe, e.g., for leptogenesis. In this

[4]The Schwarzschild geometry is described by the Schwarzschild metric, which is the most general spherically symmetric vacuum solution of the Einstein field equations.

[5]The FLRW geometry is characterized by the metric tensor, which describes a homogeneous, isotropic, expanding (or contracting) universe.

respect, $\bar{t}$ is compatible with the *nonthermal* leptogenesis period that typically dates between 10^{-26}–10^{-12}s after the Big Bang.

3.5 Financial markets and econophysics - II

Another important application of the superstatistics paradigm is in *econophysics*. As mentioned in Chapter 2, econophysics is an emerging interdisciplinary field that makes use of concepts and methods of statistical physics in order to analyze complex economical phenomena (be it financial time series, portfolio management or success and failure of companies). The validity and promise of this paradigm has been demonstrated in a number of recent publications and review papers, see e.g., Refs. [45, 179, 222, 307] and citations therein. For the reader's convenience I provide a glossary of relevant financial terms in Supplementary notes.

My take on superstatistics applications in econophysics was presented in three papers [148, 150, 155]. In particular, in Ref. [148] the so-called *stochastic adiabatic theorem* was proved, in Ref. [150] a generalization of the option pricing formula for a Gamma-superstatistics (non-Gaussian) price fluctuations was introduced and discussed, and in Ref. [155] the transitions between distinct superstatistical regimes in the context of high-frequency (minute-tick) share-price returns of seven selected companies was observed and analyzed.

Let me start first with the stochastic adiabatic theorem.

3.5.1 *Stochastic "adiabatic" theorem:*

As seen in Section 3.2 one can compute the conditional probabilities for Markovian processes by employing PI calculus. For non-quadratic Hamiltonians such computations are typically quite difficult and hence one might prefer to study such conditional probabilities directly in terms of their time-evolution equations. Such equations are called (forward) *Kramers–Moyal equations* and are of the form [269]

$$\partial_{t'} P(x', t'|x, t) \;=\; \mathbb{L} P(x', t'|x, t)\,, \tag{3.38}$$

where the Kramers–Moyal operator $\mathbb{L}$ has the expansion

$$\mathbb{L}(\partial_{t'}, x') \;=\; \sum_{n=1}^{\infty} (-\partial_{x'})^n D^{(n)}(x', t')\,. \tag{3.39}$$

Here the coefficients $D^{(n)}(x',t')$ are equal to the moments of the short-time conditional probability $P(x',t'|x,t)$, namely

$$\begin{aligned} D^{(n)}(x,t) &= \frac{1}{n!}\,\lim_{\tau\to 0}\,\frac{1}{\tau}\int_{-\infty}^{\infty} dz\,(z-x)^n\,P(z,t+\tau|x,t) \\ &= \frac{1}{n!}\,\lim_{\tau\to 0}\,\frac{1}{\tau}\int_{-\infty}^{\infty} dz\,(z-x)^n\,\langle z|e^{-\tau\hat{H}}|y\rangle\,, \end{aligned} \tag{3.40}$$

cf. e.g., [148, 269]. According to *Pawula's theorem* [249] (see also Appendix J) there is either infinite number of non-zero coefficients $D^{(n)}(x,t)$ or, if there is a finite number of them, they can be non-zero only up to $n=2$. In addition, any artificial truncation of of expansion (3.39) at $n>2$ would produce non-positive conditional probabilities. This is the basic reason why phenomenological models for $\mathbb{L}$ go typically only up to $n=2$, i.e. to Fokker–Planck operator $\mathbb{L}^{\rm FP}$ and ensuing (forward) *Fokker–Planck equation*.

Because Eq. (3.38) involves derivatives with respect to final coordinates (x',t'), it is also called *forward* Kramers–Moyal equation. To solve this equation one has to specify boundary conditions at the initial time t. Since $\hat{H}$ might contain diffusive terms, e.g. terms proportional to momenta, backward time evolution is generally not described by the same operator $\mathbb{L}$ as the forward evolution. Though the backward Kramers–Moyal equations are easy to construct (see e.g. [269]), they are less relevant in practical applications, and hence I shall deal here only with their forward counterparts.

Let me first start with the following theorem.

Theorem 3.1 ("Adiabatic" theorem). *If both* $P_\zeta(x',t'|x,t)$ *and*

$$P(x',t'|x,t) \;=\; \int d\zeta\; P_\zeta(x',t'|x,t)\,\omega(\zeta,T) \quad \textit{with} \quad T=t'-t\,,$$

are Markovian PDF's then their dynamics can be described via two (forward) Fokker–Planck equations:

$$\begin{aligned} \partial_T\,\omega(\zeta,T) &= \mathbb{L}^{\rm FP}_\omega\,\omega(\zeta,T)\,, \\ \partial_{t'}P_\zeta(x',t'|x,t) &= \mathbb{L}^{\rm FP}_\zeta P_\zeta(x',t'|x,t)\,, \end{aligned} \tag{3.41}$$

with

$$\begin{aligned} \mathbb{L}^{\rm FP}_\omega &= -\partial_\zeta\,K^{(1)}(\zeta,T) \;+\; \partial^2_\zeta\,K^{(2)}(\zeta,T)\,, \\ \mathbb{L}^{\rm FP}_\zeta &= -\partial_{x'}\,D^{(1)}_\zeta(x',t') \;+\; \partial^2_{x'}\,D^{(2)}_\zeta(x',t')\,. \end{aligned} \tag{3.42}$$

Here $K^{(1)}, D_\zeta^{(1)}$ are drift coefficients and $K^{(2)}, D_\zeta^{(2)}$ are diffusion coefficients. In addition, the previous Fokker–Planck equations are on the level of sample trajectories equivalent to two coupled Itô stochastic equations

$$\begin{aligned} dx' &= D_\zeta^{(1)}(x',t')\,dt' + \sqrt{2D_\zeta^{(2)}(x',t')}\,dW_1\,, \\ d\zeta &= K^{(1)}(\zeta,T)\,dT + \sqrt{2K^{(2)}(\zeta,T)}\,dW_2\,. \end{aligned} \tag{3.43}$$

Here $W_1(t')$ and $W_2(T)$ are Wiener processes.

Proof of the theorem is rather lengthy and I would refer the interested reader to the original paper [148]. Let me however, add few comments that would clarify the meaning of Theorem 3.1.

One way to study implications of the superstatistics PI (3.11) is to directly compute such PIs. This is, however, often very hard and aside from simple cases (cf. e.g. Ref. [113]), such computations is accessible only through perturbative methods. Often more easier is to solve associated Kramers–Moyal or Fokker–Planck equations. Since both $P_\zeta(x',t'|x,t)$ and $P(x',t'|x,t)$ satisfy the Chapman–Kolmogorov equation, it would be natural to study Kramers–Moyal or Fokker–Planck equations associated with those transition probabilitis. This might still be quite difficult task due to (typically) non-linear nature of the emergent Hamiltonian $\mathcal{H} = F(H)$. The adiabatic theorem allows to circumvent this by solving equivalent (but often simpler) problem in terms of two coupled Fokker–Planck equations (3.41). In the latter case, the drift and diffusion coefficients $K^{(1)}$ and $K^{(2)}$, respectively, can be easily computed from first two moments of the distribution $\omega(\zeta,T)$, which are often even tabulated.

Notice, further, that there are two very different characteristic time scales involved in Eqs. (3.41): t, the "internal" time scale, representing typical time scale over which $P_\zeta(x',t'|x,t)$ changes (e.g. financial data tick scale, mean collision time or time-slicing scale), and T, the "external" time, over which the "external" parameter ζ changes. Processes where $T \gg t$, as in our case, are called *adiabatic processes* — hence the adjective *adiabatic* in name of the theorem. Basic strategy for analysing an adiabatic process is first to solve the problem with the external parameters held constant, and only at the end of the calculation allow them to vary (slowly) with time. This makes the theorem useful for practical applications.

3.5.2 *Double-stochastic process in financial markets*

The previous adiabatic theorem has a direct implications in financial markets. In particular, it implies that *Gamma superstatistics* PI is closely related to the Heston stochastic volatility model. To see this more explicitly let me start with the Gamma superstatistics PI, i.e. I assume that the distribution $w(\zeta, T)$ is a Gamma distribution with parameters $b > 0$ and $c > 0$, such that

$$w(\zeta, t) \;=\; \frac{(b)^{ct}\zeta^{ct-1}}{\Gamma(ct)}\, e^{-b\zeta}\,, \qquad \zeta \in \mathbb{R}^+\,, \tag{3.44}$$

or in terms of the smearing distribution $\omega(\zeta, t)$

$$\omega(\zeta, t) \;=\; \frac{(bt)^{ct}\zeta^{ct-1}}{\Gamma(ct)}\, e^{-bt\zeta}\,, \qquad \zeta \in \mathbb{R}^+\,. \tag{3.45}$$

Since a Gamma distribution is infinitely divisible [84], it has the Lévy–Khintchine representation of the moment-generating function cf. Eq. (3.15). It can be directly checked that in this case the Lévy measure $\mu(du) = ce^{-bu}u^{-1}\mathbf{1}_{(0,+\infty)}(u)du$ and $\alpha = 0$. With this we get

$$F(p_\zeta) \;=\; \bar{\zeta}b\,\left[\log\Big(\frac{p_\zeta}{b} \;+\; 1\Big)\right], \tag{3.46}$$

where $\bar{\zeta} = c/b$ is the mean of $\omega(\zeta, t)$. In particular, for $H = \mathbf{p}^2/2$ we have

$$\mathcal{H}(p) \;=\; \bar{\zeta}b\,\left[\log\left(\frac{\mathbf{p}^2}{2b} \;+\; 1\right)\right], \tag{3.47}$$

which leads to the superstatistics PI identity

$$\begin{aligned}
\wp(x', t'|x, t) &= \int_{\mathbf{x}(t)\,=\,\mathbf{x}}^{\mathbf{x}(t')=\mathbf{x}'} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, e^{\int_t^{t'} d\tau\left[i\mathbf{p}\dot{\mathbf{x}} \;-\; \bar{\zeta}b\log\left(\mathbf{p}^2/2b+1\right)\right]} \\
&= \int_0^\infty d\zeta\, \omega(\zeta, T) \int_{\mathbf{x}(0)\,=\,\mathbf{x}}^{\mathbf{x}(T)\,=\,\mathbf{x}'} \mathcal{D}\mathbf{x}\mathcal{D}\mathbf{p}\, e^{\int_t^{t'} d\tau(i\mathbf{p}\dot{\mathbf{x}} - \zeta\mathbf{p}^2/2)} \\
&\equiv \int_0^\infty d\zeta\, \omega(\zeta, T)\, \wp_\zeta(x', t'|x, t)\,,
\end{aligned} \tag{3.48}$$

with $T = t' - t$. Since the RHS of (3.48) involves only PI with quadratic action, I can integrate it out to obtain

$$\int_0^\infty d\zeta\, \frac{(bT)^{cT}\zeta^{cT-1}}{\Gamma(cT)} \sqrt{\frac{1}{2\pi T\zeta}}\, e^{-bT\zeta} e^{-\mathbf{r}^2/2\zeta T}\,, \tag{3.49}$$

where $\mathbf{r} = \mathbf{x}' - \mathbf{x}$. After the integration I find

$$K_{1/2-cT}(2\sqrt{2b}\,|\mathbf{r}|)\,\frac{T^{-3/2}}{\sqrt{\pi}\,\Gamma(cT)}\left(\frac{2\sqrt{2b}\,T}{|\mathbf{r}|}\right)^{1/2-cT}. \tag{3.50}$$

It is quite interesting to note that in Fourier space, this coincides with a Tsallis distribution [24, 25, 150, 333].

Connection with theory of financial markets arises when I consider the free-particle Hamiltonian with drift: $p^2/2 + ip(r/\zeta - 1/2)$, see e.g., Refs. [150, 179]. Here r is the *drift constant* and ζ *volatility* (see Supplementary notes H.1). In order to find forward Fokker–Planck equations for $\wp_\zeta(x', t'|x, t)$ and $\omega(\zeta, T)$, I need to compute corresponding *drift* and *diffusion coefficients* $D_\zeta^{(1)}$ and $D_\zeta^{(2)}$ for $\wp_\zeta(x', t'|x, t)$ and similarly $K^{(1)}$ and $K^{(2)}$ for $\omega(\zeta, T)$. By employing the prescription of Eq. (3.40), the drift coefficients read

$$D_\zeta^{(1)} \;=\; \left(r - \frac{\zeta}{2}\right), \quad D_\zeta^{(2)}(x', t') \;=\; \frac{\zeta}{2}\,. \tag{3.51}$$

Similarly, the diffusion coefficients $K^{(n)}$ are

$$\begin{aligned} K^{(1)}(\zeta, T) \;&=\; \frac{1}{T}\left(\frac{c}{b} - \zeta\right) \;=\; \frac{1}{T}\left(\bar{\zeta} - \zeta\right), \\ K^{(n)}(\zeta, T) \;&=\; \frac{1}{T^n}\frac{c}{nb^n}\,, \qquad n \geq 2\,. \end{aligned} \tag{3.52}$$

With the help of adiabatic theorem I can write down the underlying Itō processes for corresponding sample trajectories, namely

$$\begin{aligned} dx' \;&=\; \left(r - \frac{\zeta}{2}\right) dt' \;+\; \sqrt{\zeta}\, dW_1 \\ d\zeta \;&=\; \frac{1}{T}\left(\bar{\zeta} - \zeta\right) dT \;+\; \frac{1}{T}\sqrt{\frac{\bar{\zeta}}{b}}\, dW_2\,. \end{aligned} \tag{3.53}$$

I can now view x' as a logarithm of a stock price S, and ζ and r as the associated variance and drift. If, in addition, I replace for large T the quantity $\sqrt{\bar{\zeta}}$ with $\sqrt{\zeta}$, the systems (3.53) reduces to

$$\begin{aligned} dS \;&=\; rS\, dt' \;+\; \sqrt{\zeta} S\, dW_1\,, \\ d\zeta \;&=\; \gamma\left(\bar{\zeta} - \zeta\right) dT + \varepsilon\sqrt{\zeta}\; dW_2\,, \end{aligned} \tag{3.54}$$

where $\gamma = 1/T$ and $\varepsilon = 1/(\sqrt{b}T)$. The system of equations (3.54) constitute Heston's stochastic volatility model [126] which is often used in financial

markets to model bond and currency options (for further reference on this model see, e.g., Ref. [126, 179] and citations therein). In this framework, the parameters $\bar{\zeta}$, γ and ε are interpreted as the long-time volatility average, the drift of the variance and the volatility of the variance, respectively.

Let us finally observe that in the large T limit $\omega(\zeta, T) = \delta(\zeta - \bar{\zeta})$ and so volatility is constant. In this case the solution of the Focker–Planck equation (3.41) with the initial condition $\wp_{\bar{\zeta}}(x', t|x, t) = \delta(x' - x)$ can be easily found as follows. I define first the short-time transition probability $\wp_{\bar{\zeta}}(x', t + \Delta t|x, t)$ where time increment Δt is small. With this I can write

$$\begin{aligned} \wp_{\bar{\zeta}}(x', t + \Delta t|x, t) &= \exp(\tau \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}})\, \wp_{\bar{\zeta}}(x', t|x, t) \\ &= \exp(\tau \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}})\, \delta(x' - x)\,. \end{aligned} \tag{3.55}$$

Now I expand $\exp(\tau \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}})$ in Δt, which gives

$$\begin{aligned} \wp_{\bar{\zeta}}(x', t + \Delta t|x, t) &= \left[1 + \Delta t \mathbb{L}_{\bar{\zeta}}^{\mathrm{FP}} + \mathcal{O}(\Delta t^2)\right] \delta(x' - x) \\ &= \left[1 - \Delta t \partial_x\, D_{\bar{\zeta}}^{(1)}(x, t) \;+\; \Delta t \partial_x^2\, D_{\bar{\zeta}}^{(2)}(x, t) + \mathcal{O}(\Delta t^2)\right] \delta(x' - x). \end{aligned} \tag{3.56}$$

On the second line I have exchanged x for x' inside of $[\ldots]$. By using the Fourier transform representation of the δ-function, I can further rewrite the RHS of (3.56) as

$$\begin{aligned} &\frac{1}{2\pi} \int_{-\infty}^{\infty} dp\, \left[1 - ip\Delta t D_{\bar{\zeta}}^{(1)}(x, t) \;-\; p^2 \Delta t D_{\bar{\zeta}}^{(2)}(x, t) + \mathcal{O}(\Delta t^2)\right] e^{ip(x' - x)} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp\, \exp\left[ip(x' - x) - ip\Delta t D_{\bar{\zeta}}^{(1)}(x, t) \;-\; p^2 \Delta t D_{\bar{\zeta}}^{(2)}(x, t)\right], \end{aligned} \tag{3.57}$$

where on the first line integration by parts was used and on the second line the contributions of the order $\mathcal{O}(\Delta t^2)$ were neglected. Integration of the Gaussian integral yields the short-time transition probability

$$\begin{aligned} &\wp_{\bar{\zeta}}(x', t + \Delta t|x, t) \\ &\quad = \frac{1}{2\sqrt{\pi \Delta t D_{\bar{\zeta}}^{(2)}(x, t)}} \exp\left\{-\frac{[(x' - x) - \Delta t D_{\bar{\zeta}}^{(1)}(x, t)]^2}{4\Delta t D_{\bar{\zeta}}^{(2)}(x, t)}\right\} \\ &\quad = \frac{1}{\sqrt{2\pi \Delta t \bar{\zeta}}} \exp\left\{-\frac{[(x' - x) - \Delta t(r - \bar{\zeta}/2)]^2}{2\Delta t \bar{\zeta}}\right\}. \end{aligned} \tag{3.58}$$

The short-time transition probability can be now used to find a final-time solution of the Fokker–Planck equation. To this end I divide time difference

T into $N+1$ small intervals of length $\Delta t = T/(N+1)$, defining $t_n = t + n\tau$ (so that $t = t_0$ and $t' = t_{N+1}$), and by repeatedly applying the Chapman–Kolmogorov equation (3.10), I get

$$\begin{aligned}\wp_{\bar{\zeta}}(x',t'|x,t) &= \int_{\mathbb{R}} dx_{_N} \cdots \int_{\mathbb{R}} dx_1\, \wp_{\bar{\zeta}}(x',t'|x_{_N},t_{_N}) \cdots \wp_{\bar{\zeta}}(x_1,t_1|x,t) \\ &= \left[\sqrt{\frac{1}{2\pi\Delta t\bar{\zeta}}} \prod_{k=1}^{N} \int_{\mathbb{R}} \left(dx_k \sqrt{\frac{1}{2\pi\Delta t\bar{\zeta}}} \right) \right] \\ &\times \exp\left\{ -\sum_{n=0}^{N} \frac{[(x_{n+1}-x_n) - \Delta t(r-\bar{\zeta}/2)]^2}{2\Delta t\bar{\zeta}} \right\}, \end{aligned} \tag{3.59}$$

which in the large N limit tends to the PI (cf. Appendix I)

$$\begin{aligned}&\int_{x(t)=x}^{x(t')=x'} \mathcal{D}x\ \exp\left\{ -\int_t^{t'} dt\, \frac{[\dot{x} - (r-\bar{\zeta}/2)]^2}{2\bar{\zeta}} \right\} \\ &= \int_{x(t)=x}^{x(t')=x'} \mathcal{D}x\mathcal{D}p \exp\left\{ -\int_t^{t'} dt\, \left[ip\dot{x} - \frac{\bar{\zeta}p^2}{2} - ip\left(r - \frac{\bar{\zeta}}{2}\right) \right] \right\} \\ &= \int_{x(t)=x}^{x(t')=x'} \mathcal{D}x\mathcal{D}p \exp\left\{ -\int_t^{t'} dt\, \left(ip\dot{x} - \bar{\zeta}H \right) \right\}, \end{aligned} \tag{3.60}$$

with $H = p^2/2 + ip(r/\zeta - 1/2)$. This PI representation precisely coincides with the prescription (3.48) where instead of a free-particle Hamiltonian I use the Hamiltonian with drift H. Since this PI has a well known solution [113], the full transition probability can be written in a closed form

$$\wp_{\bar{\zeta}}(x',t'|x,t) = \frac{1}{\sqrt{2\pi\bar{\zeta}T}} \exp\left\{ -\frac{\left[(x'-x) - \left(r - \frac{\bar{\zeta}}{2}\right) T \right]^2}{2\bar{\zeta}T} \right\}. \tag{3.61}$$

The transition probability distribution (3.61) is recognized as the risk-free member of the family of Gaussian martingale distributions for the stock price $S(t) = e^{x(t)}$, cf. e.g. Ref. [179].

Appendix G

Reprinted papers on superstatistics

[160] P. Jizba and F. Scardigli, *Special relativity induced by granular space*, Eur. Phys. J. C **73** (2013) 2491 + **highlights and news**

[149] P. Jizba and H. Kleinert, *Superstatistics approach to path integral for a relativistic particle*, Phys. Rev. D **82** (2010) 085016.

[148] P. Jizba and H. Kleinert, *Superpositions of probability distributions*, Phys. Rev. E **78** (2008) 03112.

[155] P. Jizba, J. Korbel, H. Lavička, M. Prokš, V. Svoboda and Ch. Beck, *Transitions between superstatistical regimes: Validity, breakdown and applications*, Physica A **493** (2018) 29.

Appendix H

Supplementary notes

H.1 Glossary of relevant financial terms

- **Geometric Brownian motion** *Geometric Brownian motion* is a special case of Itô process

$$dY(t) \;=\; a(Y,t)dt \;+\; b(Y,t)dW(t)\,, \tag{H.1}$$

where $W(t)$ is the Wiener process (or Brownian motion) and functions $a(Y,t)$ and $b(Y,t)$ are proportional to the random variable $Y(t)$. In particular, when $Y(t)$ is identified with the asset price $S(t)$ at time t (i.e. with the *spot price*), the geometric Brownian motion for $S(t)$ is described via the stochastic equation

$$dS(t) \;=\; \mu S(t)dt \;+\; \sigma S(t)dW(t)\,, \tag{H.2}$$

where μ denotes the *drift* (it measures the average growth rate of the asset price) and σ is the *volatility* (i.e., strength of price fluctuations — see in this glossary item *volatility*). With the help of Itô's lemma one can rewrite (H.2) as

$$d\log S(t) \;=\; \left(\mu - \frac{1}{2}\sigma^2\right) dt \;+\; \sigma dW(t)\,. \tag{H.3}$$

Thus, it is $\log S(t)$, and not the spot price $S(t)$ itself that performs a Wiener process with a (constant) drift. In other words, it is the *log-return* $r_{dt}(t+dt) \;=\; d\log S(t)$ rather than the absolute change $dS(t) = S(t+dt) - S(t)$, which is the relevant financial quantity (see in this glossary item *returns*).

- **Options** *Options* are financial derivatives (i.e., a particular type of financial instruments) that give buyers the right, but not the obligation(!), to buy or sell (depending on the type of contract they hold) an

underlying asset at an agreed-upon price and date. There are two types of options; a) *call options*, which allow the holder to buy the asset at a stated price within a specific time frame and b) *put options*, which allow the holder to sell the asset at a stated price within a specific time frame.

According to how one treats the dates on which the option may be exercised, options fall into various classes that are called *styles*. The vast majority of options are either *European* or *American* style options (or simply European or American options). These options (as well as others where the payoff is calculated similarly) are referred to as *vanilla options*. Options where the payoff is calculated differently are known as *exotic options*. For instance, European options can be exercised only at the expiration date of the option, i.e. at a single pre-defined point in time, while American options may be exercised at any time before the expiration date.

In 1973, Fisher Black, Myron Scholes and independently Robert Merton used the geometric Brownian motion to construct theory for determining the price of *European style options*. The key property of their model is that it shows that the European option has a unique price regardless of the risk of the underlying security and its expected return. This new paradigm has formed a landmark in the development of mathematical finance and currently represents a standard tool in daily capital market practice. Despite the fact that Merton and Scholes received the Nobel Prize jointly in 1997 for their work on the option pricing formula — *Black–Scholes option pricing formula*, the pricing formula itself is far from being perfect. For instance, significant limitations arise from:

- The underestimation of extreme moves in the stock, which yields heavy-tail risk.
- The assumption of instant, cost-less trading (though banks always take a manipulation fee), which yields liquidity risk.
- The assumption of a stationary process, which yields volatility risk (empirical volatility is time dependent).
- The assumption of continuous time and trading, which yields the so-called gap risk (e.g. trading hours per day are limited).

In practice there exist diverse investment strategies trying to account for these limitations. In addition, Black–Scholes pricing formula works only for European style options (e.g., no corresponding formula exist

for American options), which by themselves represent only a particular segment of the total option market.

- **Returns** In finance, *return* is a profit on an investment and it serves as a fundamental financial and business performance measure. In financial practice one recognizes various types of returns. For instance, *return on assets* or ROA, measures how much money a company earns by putting its assets to use. In other words, ROA is an indicator of how efficient or profitable a company is relative to its assets or the resources it owns or controls. *Return on invested capital*: is used to assess a company's efficiency at allocating the capital under its control to profitable investments. The return on invested capital ratio gives a sense of how well a company is using its capital to generate profits. *Return on investment* (ROI): is used to evaluate the efficiency or profitability of an investment. ROI tries to directly measure the amount of return on a particular investment, relative to the investment's cost. It can be calculated by dividing the profit earned on an investment by the cost of that investment.

  Quite generally, if $Y(t)$ is the price of a financial asset at a time t one can calculate the *net return* over a *single period* of time length (or time horizon) Δt as:

$$R_{\Delta t}(t+\Delta t) \;=\; \frac{Y(t+\Delta t)-Y(t)}{Y(t)}\,. \tag{H.4}$$

  For the analysis of the high-frequency financial data one can consider $\Delta t \ll 1$. Given a financial price record $Y(t)$ and a time lag Δt, one defines the price log-return in the interval Δt as

$$\begin{aligned} r_{\Delta t}(t+\Delta t) \;&=\; \log Y(t+\Delta t) - \log Y(t) \\ &=\; \log\left[1+R_{\Delta t}(t+\Delta t)\right]\,. \end{aligned} \tag{H.5}$$

  For high-frequency data $|R_{\Delta t}(t+\Delta t)| \ll 1$ which gives $r_{\Delta t}(t+\Delta t) \approx R_{\Delta t}(t+\Delta t)$. So, when one is interested in high-frequency data or when his/her investigations are limited to short term horizons the *log-returns* and *net returns* are approximately equal.

  One advantage of using log returns is simplicity in dealing with *multiperiod returns*. A k-period log-return is simply the sum of the

single-period log-returns since

$$\begin{aligned} r_{k\Delta t}(t+k\Delta t) &= \log\left[1+R_{k\Delta t}(t+k\Delta t)\right] \\ &= \log\left\{\prod_{j=1}^{k}[1+R_{\Delta t}(t+j\Delta t)]\right\} \\ &= \sum_{j=1}^{k}\log\left[1+R_{\Delta t}(t+j\Delta t)\right] \\ &= \sum_{j=1}^{k} r_{\Delta t}(t+j\Delta t)\,. \end{aligned} \tag{H.6}$$

This should be contrasted with a k-period *gross return*

$$\begin{aligned} 1+R_{k\Delta t}(t+k\Delta t) &= \frac{Y(t+k\Delta t)}{Y(t)} \\ &= \prod_{j=1}^{k}\left[\frac{Y(t+j\Delta t)}{Y(t+(j-1)\Delta t)}\right] \\ &= \prod_{j=1}^{k}\left[1+R_{\Delta t}(t+j\Delta t)\right]\,. \end{aligned} \tag{H.7}$$

One of the key assumption in modeling the evolution of asset prices is that returns from a single asset are mutually independent and identically distributed with normal distribution. So, log-returns rather than net/gross returns are better suited to satisfy these assumptions. E.g., sum of random normally distributed variables $r_{\Delta t}(t+j\Delta t)$ is normal while the products of normally distributed variables $1+R_{\Delta t}(t+j\Delta t)$ is not.

- **S&P500 index** There are some very important and less important stock market places around the world, on which one can buy or sell shares of companies. Most important are located in New York, London, Frankfurt and Tokyo. Particularly important stock market is the New York Stock Exchange (NYSE) where one of the oldest indexes — the S&P500 Index (Standard & Poor Composite Index) is traded. The S&P500 Index comprises portfolio of 500 different stocks: 400 industrials, 40 financial institutions, 40 utilities and 20 transportation

companies. Because of its representativeness of the U.S. stock market, diverse constituency, and the portfolio weighting methodology, the S&P500 Index enjoys broad applicability in economic models. For instance, analysis of the S&P 500 returns shows that the decorrelation time is $\approx$ 4 min. In fact, with the present computer massification the decorrelation times can be squeezed up to seconds. On the other hand, the decorrelation times of the S&P 500 volatility is $\approx$ 3 months. There exists a strong evidence that a higher momenta (namely skewness and kurtosis) have decorrelation times of order years.

- **Securities** A *security* is a financial instrument, typically any financial asset that can be traded. Term *security* broadly covers all traded financial assets and breaks such assets down into three primary categories: a) equity security (e.g. share of a company's stock) b) debt security (e.g. corporate and government bonds) and c) derivative (e.g. options).

- **Stock market** The financial markets and specially an important part of these, the stock markets, are places on which the companies and governments can buy and sell the so-called shares. Nearly every major stock market has his own index. For example, the New York Stock Exchange market index is known as NYSE Composite and covers more than 2000 stocks traded in the New York Stock Exchange (NYSE). Similarly, the Frankfurt Stock Exchange index is known as the DAX Index (Deutsche Aktien Index) and contains the 30 biggest and highest volume German company shares traded in Frankfurt. Typically there are more then one index on a single stock exchange. In New York there are traded 4 large indexes (Dow Jones, Average, S&P500 and NYSE Composite) while in Frankfurt it is 10 indexes. A large company will usually have its stock listed on many stock markets across the world.

- **Volatility** The *volatility* can be most broadly defined as a degree to which price moves. For instance, a stock with a price that fluctuates wildly-hits new highs and lows or moves erratically is said to be highly volatile. A stock that maintains a relatively stable price is considered to have a low volatility. A highly volatile stock is inherently riskier. For this reason, multiple measures of volatility are implemented in practice. By far the most popular measure of market volatility is *standard deviation*. Among another volatility measures belong e.g., *first absolute moment* when standard deviation is infinite (e.g., for heavy tailed dis-

tributions), *maximum drawdown* (i.e., the largest historical loss for an asset, measured from peak to trough, during a specific time period), *beta coefficient* (i.e., indication of the volatility of a stock, a fund, or a stock portfolio in comparison with the market as a whole — as a benchmark index is most commonly used the S&P 500 index) or *entropy* which quantifies ignorance and hence riskiness.

H.2 Some important biographies

There are numerous researchers that have substantially shaped our present understanding of *generalized statistics*. For completeness sake and in order to put my exposition into a historical context, I briefly sketch some biographies which I consider as a "well-balanced ensemble". My choice is inevitably subjective but as already stressed by Edward Gibbon [100]: "A choice of gurus is always personal.". Finally, for the reader convenience I organize the biographies alphabetically.

H.2.1 *L.E. Boltzmann (1844-1906)*

Ludwig Eduard Boltzmann was one of the major figures in the development of the atomic theory of matter. He was born in Vienna, Austria on February 20. After receiving his PhD from the University of Vienna in 1866, Boltzmann held professorships in mathematics and physics at Vienna, Graz, Munich, and Leipzig. He returned back to Vienna in 1895 where he remained till his death. In 1872 Boltzmann proposed the kinetic equation presently known as the Boltzmann kinetic equation. This equation determines the evolution of a single-particle velocity distribution for molecules in a dilute gas. The stationary solution of this equation gives the celebrated Maxwell–Boltzmann velocity distribution law on which the statistical thermodynamics of dilute gases was found. It should be stressed that already in 1866 Maxwell derived this distribution from vastly different (and simpler) arguments than Boltzmann.

Even more fundamental aspects of Boltzmann kinetic equations is that it provides a method for calculating properties of dilute gases in non-equilibrium states. Over time the Boltzmann equation has became a proto-

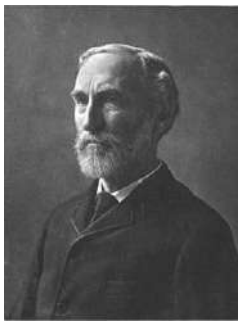
type for a general kinetic method to formulate a temporal evolution of the distribution function. The Boltzmann equation was also important mathematically, in being the first equation that described the time-evolution of a probability.

It seems more than certain that Boltzmann's atomistic ideas were central to M. Planck's later analysis of black body radiation in 1900 and A. Einstein's paper of 1905 on the nature of Brownian motion. He recognized that the assumption of equal weights of microscopic states is sufficient to build a general scheme for the statistical mechanics of equilibrium states, namely statistical thermodynamics.

Boltzmann works on the kinetic theory of gases and statistical mechanics together with his atomic hypothesis were strongly attacked many scientists including such prominent figures as E. Mach, E Zermelo and W. Ostwald. He remained misunderstood and this status que contributed badly to his mental condition which culminated in resigning his position at Vienna University in 1906. On the same year Boltzmann hanged himself while on a holiday in Duino in the north-eastern Italy.

It is rather ironic that his conclusions started to gain theoretical support already during his life by the discoveries in atomic physics that began to emerge shortly before 1900. It was recognized also soon after 1900 that fluctuation phenomena, such as Brownian motion (and later Lévy process), could be explained only through statistical concepts. Presently the most prestigious price in statistical physics is named after Boltzmann — the Boltzmann Medal.

H.2.2 *J.W. Gibbs (1839-1903)*



Josiah Willard Gibbs was born in New Haven, USA on February 11, 1839. Both his professional and personal life were confined to a single city — New Haven in Connecticut. With the exception of a three-year long visit to Europe (Berlin, Heidelberg and Paris) where he attended lectures of leading European mathematician and physicists including J. Liouville, L. Kronecker, H. Magnus, H. Helmholtz and G. Kirchoff, he studied and taught at Yale University, which is also in New Haven, and died in the same city on April 28, 1903. In 1971 Gibbs became an (unpaid) Adjunct Professor of mathematical physics at Yale. He started

to be paid after more than 10 years, and only after the John Hopkins university offered him a full professorship with a regular salary. Apparently he found it difficult to break his ties with Yale University and accepted a (financially less convenient) counter-offer from Yale.

During his scientific career was Gibbs active in thermodynamics of fluids and substances. His paper "On the Equilibrium of Heterogeneous Substances" from 1875 is still considered as a classic work in the field. It is widely believed that Gibbs was the first who systematically discussed the transformation of the fundamental equation of thermodynamics; $TdS = dU(S,V) + dW$ into equivalent forms with functional dependence on (T,V), (T,P) and (S,P). In fact, it is almost certain that it was Gibbs who coined the name *thermodynamic potential*.

As for a statistical physics, Gibbs published his famous "Elementary Principles of Statistical Mechanics" in 1902. It summarized 14 years of his research in statistical physics. There he proposed that in order to develop a macroscopic basis for thermodynamics one should instead a detailed mechanical description of the behavior of the individual constituents of matter (àla Boltzmann) consider the *statistical* behavior of system of molecules, provided one wish. To this end he introduced the concept of statistical ensemble and formulated the ensemble entropy (Gibbs's entropy) that allows to assigns probability via the maximal entropy principle to each given microstate. According to Gibbs, one should then identify the thermodynamical quantities with averages of the corresponding microscopic quantities over all systems in the ensemble. This constitutes what Gibbs called *ensemble average*.

Similarly as A. Einstein's "Relativity, the special & the general theory" or P.A.M. Dirac's "The principles of Quantum Mechanics", Gibbs' book has basically no reference to the literature. In fact, apart from his book Gibbs published no paper on the subject of Statistical Mechanics.

Gibbs name is connected with many concepts in Thermodynamics and Statistical Physics, including Gibbs paradox, Gibbs free energy, Gibbs equation, Gibbs–Duhem equation, Gibbs factor or Gibbs entropy. Among others, Gibbs also coined the term *statistical mechanics*. This apparently happened during an oral presentation of one of his papers before the American Association for the Advancement of Science in 1884.

H.2.3 *E.T. Jaynes (1922-1998)*

Among prominent researchers exploring the connection between the maximal entropy principle (MaxEnt) and Shannon's information theory was indisputably Edwin Thompson Jaynes. E.T. Jaynes was born in Waterloo, Iowa, USA on July 5, 1922. He received his B.A. in Physics from the University of Iowa in 1942 and graduated at UC-Berkeley. In 1947 he transferred to Princeton University, where he completed a Ph.D. thesis in 1950 on ferroelectricity under the supervision of Eugene Wigner. In 1960 he accepted a senior faculty appointment at the University of Washington, St. Louis, where, with few interruptions, he remained for the rest of his life.

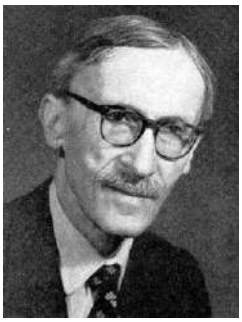
Jaynes's major contributions lie essentially in four fields; applied classical electrodynamics, information theory and statistical mechanics, quantum optics and neoclassical radiation theory, and probability theory and statistical inference. Much of his work was in one way or another controversial and in some cases it remains so till today. In 1957 he published his first two papers interpreting statistical mechanics in terms of information theory. It is true that the relationship between entropy and information was already explored before Jaynes, e.g., by Szilárd, Schrödinger, Wiener or Brillouin. In particular, it was Brillouin who, while re-examining the problem of Maxwell's demon, introduced in 1952 the notion of "negentropy". Jaynes, however, rightly pointed out that Shannon's measure of information is simply the "measure of the uncertainty inherent in a preassigned probability scheme" and *per se* it has nothing to do with thermodynamic entropy, except in the cases in which the probability distribution is known, or proven to be, "(grand-)canonical". At the same time he realized that in practice one should always adopt for the statistical description the distribution of maximal Shannon's entropy among all those which satisfy experimentally given constraints. In other words, one should always choose the least biased (most noncommittal) distribution permitted by the circumstances. The Gibbs probability distribution can be then derived from a principle of maximum entropy subject to constraints representing prior knowledge (e.g., when we know the average energy). According to Jaynes, the statistical mechanics can thus viewed as a form of statistical inference based partial information, rather than a physical theory.

Like Einstein, he did not like the Copenhagen interpretation of quantum mechanics which he viewed as an incursion of mysticism into science. Out of this line of dissatisfaction arose a work on neoclassical radiation theory, published together with Fred Cummings in 1963. This theory, known as Jaynes–Cummings model is probably his most cited and most notorious paper. In 1975 he became the Wayman Crow Distinguished Professor of Physics at Washington University in St. Louis. Jaynes died in St Louis, Missouri, USA on April 30, 1998.

H.2.4 *A. Y. Khinchin (1894-1959)*

Aleksandr Yakovlevich Khinchin (or Khintchine) was born on July 19, 1894 in the village Kondrovo (in Kaluga region) in Russian Empire. He obtained his secondary school education both in Russia (in Kaluga) and in Switzerland (in Zürich). His university study was done at Moscow State University under supervision of a famous Russian mathematician Nicolai. N. Luzin. Khinchin graduated from the university in 1916 and in 1922 he returned back to MSU as a full professor of mathematics. He stayed at MSU until his death. Khinchin's work focuses mainly on probability theory and number theory. He became one of the founders of modern probability theory. He embraced Kolmogorov's axiomatic approach and discovered the law of the iterated logarithm in 1924 and achieved important results in the field of limit theorems. Togetehr with P.P. Lévy, he was also one of the founders of theory of stationary processes. Despite his mathematical upbringing, he was extremely well versed in such subjects as statistical physics, quantum mechanics and information theory. He published two key (and influential) books: "Mathematical foundations of information theory" and "Mathematical foundations of statistical mechanics" where he used the methods of probability theory, information theory, queuing theory and mathematical analysis in rather unusual but very inspiring way. With his name are connected such concepts as Wiener–Khinchin theorem, Lévy–Khinchin formula for infinitely divisible random variables, Lévy–Khinchin formula for stable processes or Khinchin's constant. Khinchin died in Moscow on November 18, 1959.

H.2.5 *P.P. Lévy (1886-1971)*



Paul Pierre Lévy was born on September 15, 1886 in Paris. He came from a family of mathematicians. His father was a teacher at the prestigious École Polytechnique and his grandfather was a professor of mathematics. He attended the Lycée Saint Louis in Paris where he excelled in number of subject including, chemistry, physics and mathematics. After the secondary school he entered the École Polytechnique and in the age of 19 (while still undergraduate) he published his first paper on semi-convergent series. He finished his PhD under supervision of Hadamard in 1912. In 1913 he became professor at École des Mines de Saint-Étienne in Paris and then in 1920 a professor of analysis at the École Polytechnique in Paris where he remained until he retired in 1959. Among his students were, e.g., Benoit Mandelbrot. In 1919 was Lévy asked to deliver three lectures at the École Polytechnique titled "Notions of calculus of probabilities and the role of Gaussian law in the theory of errors." These changed both Lévy's life and probability theory. He discovered the class of probability distributions known as "stable distributions" (the Lévy–Khintchine formula) and proved the generalized version of the Central Limit Theorem for independent variables with infinite variance. Independently from the Soviet mathematicians Kolmogorov, Gnedenko and Khinchin, he discovered the major part of what is presently known as the theory of stochastic processes. Some of Lévy's works were originally considered as unrealistic or even obscure. Among theses were, e.g., works on processes with infinite variance (known as stable processes) or on local times of stochastic processes which turned out to be extremely useful in a number of contexts, including condensed matter physics, statistical physics and finance. In 1953 he started to used Hausdorff measure to study Brownian paths and become a pioneer of fractals, which were in 1970's popularized by his student Benoit Mandelbrot.

In 1963 was Lévy elected as an honorary member of the London Mathematical Society. In the following year he was elected to the Académie des Sciences. Lévy died on December 15, 1971 in Paris.

H.2.6 *B. Mandelbrot (1924-2010)*

Benoît Mandelbrot was born on November 20, 1924 in Warsaw, Poland into a Jewish family from Lithuania. His mother was a dentist and his father worked in the clothing business. The family fled to Paris in 1936, in time to escape Hitler's advances. Once in Paris, he was mentored by his bright mathematician uncle Szolem Mandelbrojt (who was also an early member of the Bourbaki group). After the fall of France his family fled Paris, taking refuge in the south of France (Vichy Zone libre) before the country was liberated in summer 1944. After the war he studied at the prestigious École Polytechnique in Paris, and later at the California Institute of Technology, Princeton (where he was the last postdoc of the great mathematician John von Neumann) and M.I.T. At M.I.T. he got to know the young Noam Chomsky with whom he often debated linguistic and philosophy. Mandelbrot ultimately settled at I.B.M.' Thomas J. Watson Research Center, where he stayed for 35 years. He took frequent leaves, teaching at many universities, including Harvard and Yale. In the latter he was the Sterling professor for 17 years (though he got tenure at Yale only when he was 75).

Mandebrot's life was irrevocably changed when his uncle Szolem introduced him to Zipf's Law, i.e. the law that was originally invented to deal with the frequencies of words in various languages. Mandelbrot observed that Zipf's law led to some counterintuitive and universal results that could only be explained by non-standard distributions (heavy-tailed distributions); this was when he discovered the high prevalence of what many had previous considered to be "rare" events. His work in this area as well as some preliminary work in economics led him to a highly productive position at I.B.M. It was while at I.B.M. that Mandelbrot discovered what he is most famous for — *fractals* and *fractal geometry*. The concept of fractals arose originally from analyzing price and market fluctuations. Mandelbrot was impressed by the striking similarity of disparate price and income time series and realized that the equilibrium model that economists were relying upon for decades was of little use in analyzing real world jumps which tended to be much more numerous than the standard CLT would indicate.

Today fractals as well as their sophisticated version — *multifractals* (invented by Madelbrot in early 1970's), are known to manifest themselves in a

staggering range of phenomena in physics, engineering, arts, medicine (e.g., vessels, lungs, and brains are fractal), biology, etc. Surprisingly enough, Mandelbrot was long ignored by the very field he started in — economics, despite the fact the he proved already in the 1960's that financial theories vastly underestimate market risk. His approach was not adopted by economists partly because they found it difficult to use and partly because the field was populated by "established" ideas from equilibrium models. It was only in the 1980s that his insights became accepted into the mainstream, and the global recession in 2008 and the shocks to the economy have soundly validated his (multi-)fractal fluctuation models. He died on October 14, 2010 in hospice in Cambridge, Massachusetts, USA.

H.2.7 *A. Rényi (1921-1970)*

Alfréd Rényi belongs together with Pál Erdös and John von Neumann among the most prominent Hungarian-born world mathematicians. He was born on March 20, 1921 in Budapest, Hungary to a Jewish family. In 1940 he was admitted to the University of Budapest to study mathematics and physics. After graduation in 1944 he was drafted to forced labor service, escaped, and completed his Ph.D. in 1947 at the University of Szeged, under the advisement of the famous mathematician Frigyes Riesz. After a brief assistant professorship and Privatdozentship at Budapest University, he was appointed in 1949 as Extraordinary Professor at the University of Debrecen and held this position till 1950, when he founded the Mathematics Research Institute of the Hungarian Academy of Sciences, now bearing his name. From 1952, he also headed the Department of Probability and Mathematical Statistics of the Eötvös Loránd University in Budapest. Rényi counts as a founder of Hungarian Probability Theory School — now worldwide renown.

During his short life he was exceptionally prolific. He produced joint papers with 36 co-authors on variety of mathematical subjects. For instance, with Erdös alone he wrote 32 papers. As a visiting professor he spent a substantial time period at number of universities, mostly in the USA. Between 1965 and 1969, he was a member, and then vice president of the International Statistical Institute

In 1961 he introduced the one-parametric class of information measures

of order α — presently known as Rényi's entropies, which give an important generalization of the Shannon entropy and the Kullback–Leibler divergence.

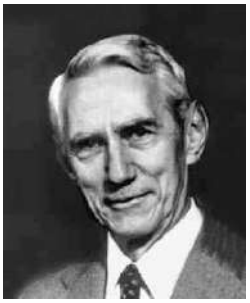
Rényi name is connected with many concepts in information theory probability theory and statistics including Rényi entropies, Rényi–Ulam game, Erdös–Rényi phase transition (also known as percolation transition), etc. He died at the lung cancer at the age of 49 in 1970.

H.2.8 *E. Schrödinger (1887-1961)*

His father ran a small linoleum factory in Erdberg, Austria. He studied mathematics and physics at the University of Vienna, receiving a doctorate on the conduction of electricity on the surface of insulators in moist air. He then undertook voluntary military service until he obtained an assistantship at Vienna. During World War I he saw active service in Italy and Hungary, receiving a citation for outstanding service, but in 1917 he returned to Vienna, where he worked on color theory, and moved to Jena and then Zürich in 1921. He began to study atomic structure and quantum statistics and was influenced by L. de Broglie's thesis. In 1926 he published his revolutionary wave mechanics including his great discovery, Schrödinger's wave equation. According to his own admission "... *this work was a result of my dissatisfaction with the quantum condition in the Bohr's theory of the atom and my believe that atomic spectra should really be determined by some kind of eigenvalue problem.*" In 1927 he moved to Berlin as Planck's successor at the University of Berlin (currently the Humboldt University). While in Berlin, he published his famous derivation od the Heisenberg's uncertainty relation based on the Schwarz inequality. In 1929 he was elected as the youngest member of the Prussian Academy of Sciences. This was at that time the most prestigious scientific honor. Due to the rise of the Nazi Party (NSDAP) and ensuing racial laws excluding Jews and other 'undesirables' from holding governmental (including academic) posts Shrödinger resigned his membership of the Prussian Academy and later in 1933 also his post at Berlin University. After this moved from place to place (including Oxford University, Princeton University and the University of Graz), finally taking appointment at the Dublin Institute for Advanced Studies in 1939. In 1933 he was awarded jointly with Paul Dirac the Nobel Prize in Physics and

in 1935 (while at Oxford University) his famous Schrödinger's cat paradox appeared. He published work on a unified field theories, but was devastated by Einstein's criticism of this work. He also wrote various books including "What is life?" and "Statistical thermodynamics", before returning to Vienna in 1956. There Schrödinger was appointed to a professorial post specially created for him at the University of Vienna. Apart from his work on quantum mechanics and general relativity he dedicated also a substantial part of his work to statistical physics and thermodynamics. Known are particularly his influential papers on the statistical meaning of entropy (quantifying the concepts of order and disorder) and quantum statistical physics (e.g., Schrödinger–Park paradox). He was also first to discuss the concept of negentropy and its relation to genetic information stored in molecules (later named as the DNA molecule). He died at the age of 73 years on January 4, 1961 in Vienna. The cause of death was simply given as old age but most likely it was tuberculosis.

H.2.9 *C.E. Shannon (1916-2001)*



Claude Elwood Shannon was born on April 30, 1916 in Petoskey, Michigan, USA. His father was mathematically skilful and worked as a probate judge in Gaylord, Michigan. His mother was a language teacher in Gaylord. From the early childhood he had remarkable mechanical aptitude and liked mathematical puzzles. In 1932 Shannon started his undergraduate studies at the University of Michigan. In 1936 he was awarded a first degree in electrical engineering and another in mathematics. He maintained his love for both subjects throughout his life. In 1936 he joined Department of Electrical Engineering at the M.I.T. as a research assistant. After two years, he transferred to the Department of Mathematics, where he wrote a brilliant master's thesis entitled "An analysis of relay and switching circuits". In this work he showed that the Boolean binary arithmetic could be represented by electrical circuits. As a result he was able to develop mathematical techniques for building a network of switches and relays to realize a specific logical function.

In 1940 he completed his Ph.D. and was awarded a National Research Fellowship to work at the Institute for Advanced Study (IAS) in Princeton, New Jersey. His research line at IAS was strongly influenced by John von

Neumann, whose many interests included computing, automata and game theory. In 1940 was Shannon recruited as a research mathematicians for the Bell Telephone Laboratories in Manhattan. Among his colleagues were at that time John Bardeen, Walter H. Brattain and William B. Shockley who received the Nobel Prize in Physics for the invention of the transistor.

In 1948 he published his most influential paper "A Mathematical Theory of Communication". The paper is often considered as the most important engineering paper ever written. There he laid down mathematical foundations of a new mathematical discipline — communication theory (now called information theory). As stressed by the famous Russian mathematician A.I. Khinchin: *Rarely does it happen in mathematics that a new discipline achieves the character of a mature and developed scientific theory in the first investigation devoted to it. .... so it was with information theory after the work of Shannon.*

The axiomatic style of the paper paper not only allowed to provide an operational meaning to hitherto vague notion of *information* but allowed Shannon to formulate his (now famous) coding theorem. In particular, in his *source* coding theorem Shannon shows that it is impossible to compress the source data so that the expected length of a source code is less than the Shannon entropy of the source, without loosing some part of the original source information. Interestingly enough, Shannon entropy is identical in its form (apart from a scaling factor) to Gibb's expression for thermodynamic entropy. This had a wide repercussions for a development of statistical physics and information technology in the second half on 20th century.

Apart from information theory, Shannon was scientifically prolific also in the game theory and cybernetics. His pioneering work on computer chess and on mechanical "mouse" (Theseus) that learn the path through a maze are among the earliest contributions to the field of artificial intelligence. He was also skilful electro-engineer and to the general public he was well known for numerous ingenious machines and gadgets he invented and built.

In 1956 Shannon came to M.I.T. to join its faculty as the Donner Professor of Science. At the M.I.T. he worked in the Research Laboratory of Electronics from which he formally retired in 1978. Interestingly enough, Shannon's publication activity essentially came to an end in 1967 when he was just over 50 years. By the end of 1990's he developed Alzheimer's disease and died at the age of 84 years on February 24, 2001 in a nursing home in Medford, Massachusetts, USA.

H.2.10 *C. Tsallis (1943-∗)*

Constantino Tsallis is a Brazilian physicist of Greek origin. He was born on November 5, 1943 in Athens. He grew up in Argentina, where he studied physics at Instituto Balseiro, in Bariloche. In 1974, he received a Doctorat d'État ès Sciences Physiques degree from Université Paris Sud-Orsay under supervision of Andre Guinier. His scientific interests has focused since on foundational issues of statistical physics. In 1988 he published in Journal of Statistical Physics an influential paper in which he proposed a generalization of Boltzmann–Gibbs entropy and ensuing modifications of conventional statistical physics. It was later on found that his generalized entropy — presently known as Tsallis entropy, can be identified with Havrda–Charvát's information measure (or structural α entropy). Tsallis put the Havrda–Charvát's measure "on the map" by stressing that it represents a natural entropy (or complexity measure) for a specific class of anomalous systems that are characterised by non-ergodicity and for a number of systems with metastable states. Tsallis seminal paper became a fertile ground for several further investigations in the area of non-equilibrium statistical physics. To date, there is in a number of practical applications of Tsallis thermostatistics in the areas of critical phenomena, chaos and nonlinear dynamics, economics, cognitive psychology, immunology, population evolution, etc.

Presently Tsallis holds a joint appointment with the National Academy of Sciences of Brazil and the Santa Fe Institute, New Mexico.

Appendix I

Crash course on path integral

I.1 Quantum mechanics — Feynman path integral

For logical consistency I will briefly sketch here some essentials of Feynman's path integral (PI). For more detailed exposition I would relegate the reader to Refs. [92, 179, 335]. For simplicity's sake I work with a bosonic particle in one dimension. Generalization to more dimensions is quite straightforward. Fermionic PI which is typically formulated in terms of Grassmann variables will not be discussed here as it is not needed in the main text.

In its essence, Feynman's PI is a particular representation for the transition amplitude $\langle x_f, t_f | x_i, t_i \rangle$. Simplest way how to introduce PI (also the way I follow here) is to assume validity of conventional canonical quantization and derive PI from it. An alternative viewpoint where PI is introduced heuristically is presented, e.g. in Ref [92].

Let me start with the Heisenberg-picture resolution of unity

$$\mathbb{1} \;=\; \int_{\mathbb{R}} dx \; |x,t\rangle\langle x,t| \,, \tag{I.1}$$

which is valid for any time t. I now partition the time interval $[t_i, t_f]$ into $N+1$ equidistant pieces Δt (see Fig. I.1) by writing $t_f - t_i = (N+1)\Delta t$. Consequently I get that

$$\begin{aligned} \langle x_f, t_f | x_i, t_i \rangle \;=\; & \left(\int_{\mathbb{R}^N} \prod_{k=1}^{N} dx_k \right) \; \langle x_f, t_f | x_N, t_f - \Delta t \rangle \\ & \times \langle x_N, t_f - \Delta t | x_{N-1}, t_f - 2\Delta t \rangle \\ & \times \langle x_{N-1}, t_f - 2\Delta t | x_{N-2}, t_f - 3\Delta t \rangle \\ & \qquad \vdots \\ & \times \langle x_1, t_i + \Delta t | x_i, t_i \rangle \,. \end{aligned} \tag{I.2}$$

For a future reference it will be convenient to formally set $t_0 = t_i$ and

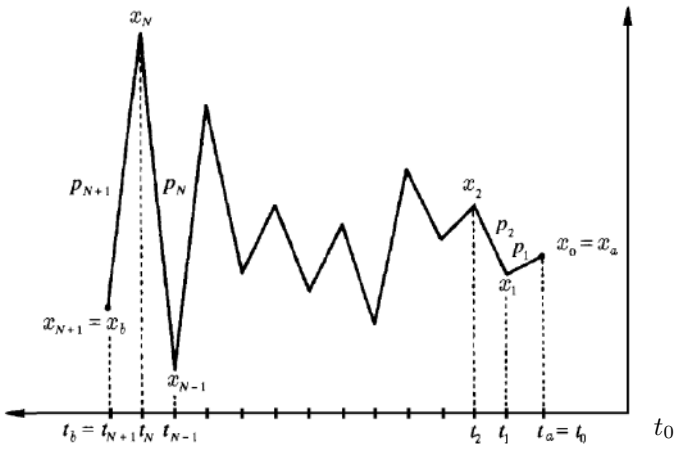


Fig. I.1: Sliced time interval $[t_b, t_a]$ with representative trajectory. Note that momenta p_{i+1} in each interval $\Delta t = t_{i+1} - t_i$ are constant and change discontinuously from one interval Δt to another. See also Eqs. (I.8)-(I.9). The right-to-left quantum mechanical convention is employed. Figure is reproduced from Ref. [179].

$t_{N+1} = t_f$.

To proceed, I utilize the trick that consists in rewriting the infinitesimal-time transition amplitudes in terms of the Schrödinger picture base vectors $|x_k\rangle$, i.e.

$$\langle x_j, t_j | x_{j-1}, t_{j-1} \rangle \;=\; \langle x_j | T \left[\exp\left(-i \int_{t_{j-1}}^{t_j} dt \; \hat{H}(t) \right) \right] | x_{j-1} \rangle \,, \qquad \text{(I.3)}$$

where the time ordering prescription T must be enfoced when $\hat{H}$ is explicitly time dependent (e.g., when it includes time dependent external fields).

In connection with (I.3) I should emphasize that the Schrödinger-picture *base vectors* are time independent in contrast with Schrödinger-picture *state vectors* that are, of course, time dependent (opposite statement holds for

Heisenberg picture). By using the expansion

$$T\left[\exp\left(-i\int_{t_{j-1}}^{t_j} dt\ \hat{H}(t)\right)\right]$$
$$= 1 - i\int_{t_{j-1}}^{t_j} dt_1\ \hat{H}(t_1) - \int_{t_{j-1}}^{t_j} dt_1\ \hat{H}(t_1)\int_{t_{j-1}}^{t_1} dt_2\ \hat{H}(t_2)$$
$$+ i\int_{t_{j-1}}^{t_j} dt_1\ \hat{H}(t_1)\int_{t_{j-1}}^{t_1} dt_2\ \hat{H}(t_2)\int_{t_{j-1}}^{t_2} dt_3\ \hat{H}(t_3)\dots, \qquad \text{(I.4)}$$

For a very small Δt I can employ the resolution of unity for the *momentum* base vectors (in Schrödinger picture) and write

$$\langle x_j, t_j | x_{j-1}, t_{j-1}\rangle$$
$$\simeq \langle x_j|\left(1 - i\int_{t_{j-1}}^{t_j} dt\ \hat{H}(t)\right)|x_{j-1}\rangle$$
$$\simeq \int_{\mathbb{R}} dp_j \langle x_j|p_j\rangle\ [1 - i\ H_{cl}(p_j, x_{j-1}, t_{j-1})\Delta t]\,\langle p_j|x_{j-1}\rangle$$
$$\simeq \int_{\mathbb{R}} \frac{dp_j}{2\pi}\ e^{ip_j(x_j - x_{j-1}) - iH_{cl}(p_j, x_{j-1}, t_{j-1})\Delta t}. \qquad \text{(I.5)}$$

Here I have defined

$$H_{cl}(p_j, x_{j-1}, t_{j-1}) = \langle p_j|\hat{H}(t_{j-1})|x_{j-1}\rangle, \qquad \text{(I.6)}$$

and $t_j = t_i + j\Delta t$. Note that H_{cl} coincides with the classical Hamiltonian if $\hat{H}$ is in the "px-ordered form" in which all $\hat{p}$-operators stand left of all $\hat{x}$-operators. If $\hat{H}$ is not in the px-ordered form then there are in H_{cl} correction of order $\mathcal{O}(\hbar)$ in comparison with the classical Hamiltonian.

By inserting (I.5) into the formula (I.2) we get exponential including the sum of Hamiltonians and term

$$\exp\left[i\sum_{k=1}^{N+1} p_k\left(\frac{x_k - x_{k-1}}{\Delta t}\right)\Delta t\right], \qquad \text{(I.7)}$$

which in the $N \to \infty$ limit ("continuous" time limit) goes over to

$$\exp\left(i\int_{t_i}^{t_f} dt\ p(t)\dot{x}(t)\right).$$

Similarly

$$\exp\left(-i\sum_{k=1}^{N+1} H_{cl}(p_k, x_{k-1}, t_{k-1})\Delta t\right) \overset{N\to\infty}{\to} \exp\left(-i\int_{t_i}^{t_f} dt\ H_{cl}(p(t), x(t), t)\right).$$

If, in addition, one adopts the formal notation

$$\lim_{N\to\infty}\left[\prod_{k=1}^{N}\int_{\mathbb{R}^2}\left(\frac{dp_k dx_k}{2\pi}\right)\right]\int_{\mathbb{R}}\frac{dp_{N+1}}{2\pi} = \int_{x(t_i)=x_i}^{x(t_f)=x_f}\mathcal{D}p\mathcal{D}x\,, \qquad \text{(I.8)}$$

it is possible to rewrite the amplitude of transition (I.2) in the so-called phase-space path-integral representation as

$$\langle x_f, t_f|x_i, t_i\rangle = \int_{x(t_i)=x_i}^{x(t_f)=x_f}\mathcal{D}x\mathcal{D}p\ \exp\left[i\int_{t_i}^{t_f} dt\ (p\dot{x} - H_{cl})\right]. \qquad \text{(I.9)}$$

Note that since p and x are not related in (I.9), the expression $p(t)\dot{x}(t) - H_{cl}(p(t), x(t), t)$ is not the Lagrangian despite its suggestive form [see note after Eq. (I.14)]. At this point one can ask: Given PI (I.9) with H_{cl} being the conventional the classical Hamiltonian, what is the most general (time-independent) Hamiltonian for which (I.9) holds. The answer is the "Weyl-ordered" Hamiltonian $\hat{H}_W$. So, in particular in (I.9) we should use

$$\langle x_f, t_f|x_i, t_i\rangle = \langle x_f|e^{-i(t_f-t_i)\hat{H}_W}|t_i\rangle\,. \qquad \text{(I.10)}$$

The proof of this statement can be found, e.g., in Ref. [20].

The phase-space PI (I.9) was firstly derived by Feynman in Ref. [88]. It heuristically represents a sum over trajectories in phase space where each trajectory carries the phase factor proportional to the classical action evaluated along respective trajectory. The concept of the PI phase-space trajectories is, however, quite troublesome from mathematical (as well as conceptual) point of view as such trajectories are discontinuous almost everywhere (see Fig. I.1). There is an extensive literature on this topic. The interested reader may consult, e.g. Refs. [54, 76, 177, 282, 300]. A mathematically cleaner alternative to phase-space PI is the so-called coherent-state PI introduced by Klauder [177].

When $H_{cl} = \frac{p^2}{2m} + V(x)$ with V being momentum independent, one can reduce the phase-space PI to the configuration-space PI. The actual argument goes as follows. I first rewrite (I.9) as

$$\int_{x(t_i)=x_i}^{x(t_f)=x_f}\mathcal{D}x\left\{e^{-i\int_{t_i}^{t_f} dt\, V(x)}\int\mathcal{D}p\ \exp\left[i\int_{t_i}^{t_f} dt\ \left(p\dot{x} - \frac{p^2}{2m}\right)\right]\right\}.$$

The PI within $\{\ldots\}$ can be further recast to the form

$$\begin{aligned}
&\int \mathcal{D}p \, \exp\left[i\int_{t_i}^{t_f} dt \, \left(p\dot{x} - \frac{p^2}{2m}\right)\right] \\
&= \lim_{N\to\infty} \int_{\mathbb{R}^{N+1}} \left(\prod_{k=1}^{N+1} \frac{dp_k}{2\pi}\right) \exp\left[i\Delta t \sum_{i=1}^{N+1} \left(p_i \frac{(x_i - x_{i-1})}{\Delta t} - \frac{p_i^2}{2m}\right)\right] \\
&= \lim_{N\to\infty} \prod_{k=1}^{N+1} \int_{\mathbb{R}} \frac{dp}{2\pi} \, \exp\left[i\Delta t \, \left(p\frac{(x_i - x_{i-1})}{\Delta t} - \frac{p^2}{2m}\right)\right] \\
&= \lim_{N\to\infty} \prod_{k=1}^{N+1} \sqrt{\frac{m}{i2\pi\Delta t}} \, \exp\left[i\Delta t \sum_{j=1}^{N+1} \frac{m(x_j - x_{j-1})^2}{2\Delta t^2}\right] \\
&= \mathcal{N} \exp\left[i\int_{t_i}^{t_f} dt \, \frac{m}{2}\, \dot{x}^2\right], \qquad \text{(I.11)}
\end{aligned}$$

where $\mathcal{N}$ is the (infinite) constant factor that can be assimilated to the the definition of $\mathcal{D}x$. On the fourth line I have used the *Fresnel integral*

$$\int_{\mathbb{R}} dx \, e^{iax^2} = \sqrt{\frac{\pi}{|a|}} \, e^{i \,\mathrm{sign}(a)\, \pi/4}, \qquad \text{for all } a \in \mathbb{R}\,. \qquad \text{(I.12)}$$

Identity (I.11) represents a special case of the so-called *Hubbard–Stratonovich transformation*. Consequently, I can formulate the transition amplitude $\langle x_f, t_f | x_i, t_i\rangle$ in an alternative form

$$\langle x_f, t_f | x_i, t_i\rangle = \int_{x(t_i)=x_i}^{x(t_f)=x_f} \mathcal{D}x \, \exp\left(i\int_{t_i}^{t_f} dt \, L(\dot{x}, x)\right). \qquad \text{(I.13)}$$

Here, $L = m\dot{x}^2/2 - V(x)$ is the conventional configuration-space Lagrangian and

$$\lim_{N\to\infty} \left[\sqrt{\frac{m}{i2\pi\Delta t}} \prod_{k=1}^{N} \int_{\mathbb{R}} \left(dx_k \sqrt{\frac{m}{i2\pi\Delta t}}\right)\right] = \int_{x(t_i)=x_i}^{x(t_f)=x_f} \mathcal{D}x\,, \qquad \text{(I.14)}$$

defines the so-called Feynman's "measure", which in fact is not a proper measure in a mathematical sense [see the comment after Eq. (I.23)].

The integration over $x_1, \ldots, x_N$ can be interpreted as summing over all possible broken line paths connecting x_i and x_f. Since any continuous path can be approximated by a broken-line path, the limit $N \to \infty$ can be viewed as a passage to a sum over all continuos paths in the configuration space.

In connection with (I.13) it should be noted that the Hubbard–Stratonovich transformation plays on the level of PI analogous role as the Legendre transform in the classical physics. In fact, both concepts will coincide when the stationary phase (or WKB) approximation is applied to the PI (I.11). In such a case only classical (on-shell) configurations for p will contribute. In fact, the stationary phase leads to the relation $p_{st} = m\dot{x}/2 = \partial L/\partial\dot{x}$, which is the required relation when passing from Hamiltonian to Lagrangian picture.

The form (I.13) often serves as a starting point for discussions concerning the classical or semiclassical limits of quantum theory. The PI (I.13) also serves as an excellent tool for setting up perturbation treatment [179]. Although ensuing technical and mathematical issues are both important, intriguing and interesting they lay outside the scope of this work and thus I shall not dwell on them. The corresponding technical aspects are well covered, e.g., in Refs. [4, 54, 104, 270].

I.2 Statistical physics and Euclidean PI

Let me now briefly outline an important connection between statistical physics and quantum mechanics that can be exemplified in terms of PIs. To this end I consider the matrix element

$$\rho(x_i, x_f, \beta) \;\equiv\; \langle x_f|e^{-\beta\hat{H}}|x_i\rangle\,, \tag{I.15}$$

of the Gibbs operator $e^{-\beta\hat{H}}$, where $\beta = 1/(k_B T)$ is the inverse temperature and k_B is the Boltzmann constant. The matrix $\rho(x_i, x_f, \beta)$, known also as the *Bloch density matrix*, is a fundamental object in quantum statistical physics, as the expectation value of an operator $\hat{O}$ at the temperature T can be written in the form

$$\langle\hat{O}\rangle \;=\; \frac{1}{Z}\int_{\mathbb{R}}\int_{\mathbb{R}} dx_i dx_f\; \rho(x_i, x_f, \beta)\,\langle x_f|\hat{O}|x_i\rangle\,, \tag{I.16}$$

where $Z = \int_{\mathbb{R}} dx\,\rho(x, x, \beta)$ is the partition function of the system.

The matrix element (I.15) may be thought as the matrix element of the *imaginary time* evolution operator

$$\mathcal{U}(x_f, \tau_f; x_i, \tau_i) \;=\; \langle x_f|e^{-(\tau_f-\tau_i)\hat{H}}|x_i\rangle\,, \tag{I.17}$$

evaluated for the interval $\tau_f - \tau_i = \beta$. Given the observation that the essence of PI is the subdivision of the time interval into sufficiently small intervals so that in the expansion (I.4) one can neglect higher-order terms

(safe for the liner one), all the steps in the derivation of the real-time PI from previous subsection can be repeated for the case of imaginary time. So, by combining (I.2) and (I.5) with (I.17) I can write

$$\begin{aligned}\rho(x_i,x_f,\beta) &\equiv \mathcal{U}(x_f,\tau_f;x_i,\tau_i)\\ &= \lim_{N\to\infty}\left[\prod_{k=1}^{N}\int_{\mathbb{R}^2}\left(\frac{dp_k dx_k}{2\pi}\right)\right]\int_{\mathbb{R}}\frac{dp_{N+1}}{2\pi}\\ &\quad\times e^{\sum_{j=1}^{N+1}[ip_j(x_j-x_{j-1})-H_{cl}(p_j,x_{j-1})\Delta\tau]}\\ &= \int_{x(\tau_i)=x_i}^{x(\tau_f)=x_f}\mathcal{D}p\mathcal{D}x\,\exp\left[\int_{\tau_i}^{\tau_f}dt\;(ip\dot{x}-H_{cl})\right]. \quad \text{(I.18)}\end{aligned}$$

Note in particular, that there is "i" factor in front of $p\dot{x}$ but not in front of H_{cl}. By using the fact that $\mathcal{U}(x_f,\tau_f;x_i,\tau_i)=\mathcal{U}(x_f,\tau_f-\tau_i;x_i,0)$ one can also rewrite (I.18) as

$$\rho(x_i,x_f,\beta) = \int_{x(0)=x_i}^{x(\beta)=x_f}\mathcal{D}p\mathcal{D}x\,\exp\left[\int_0^{\beta}dt\;(ip\dot{x}-H_{cl})\right]. \quad \text{(I.19)}$$

In cases when $H_{cl}=\frac{p^2}{2m}+V(x)$ with V being momentum independent, one can easily integrate over p. By employing the Hubbard–Stratonovich transformation together with the Gaussian integral, one obtains the configuration-space PI representation of the density matrix in the form

$$\rho(x_i,x_f,\beta) = \int_{x(0)=x_i}^{x(\beta)=x_f}\mathcal{D}x\;\exp\left(-\int_{t_i}^{t_f}dt\;H(\dot{x},x)\right), \quad \text{(I.20)}$$

where $H(\dot{x},x)=\frac{m}{2}\dot{x}^2+V(x)$ can be identified with the classical Hamiltonian function, in which the momentum p is substituted for $m\dot{x}$. In addition, Feynman's "measure" (I.14) changes to the *Euclidean measure*

$$\lim_{N\to\infty}\left[\sqrt{\frac{m}{2\pi\Delta\tau}}\prod_{k=1}^{N}\int_{\mathbb{R}}\left(dx_k\sqrt{\frac{m}{2\pi\Delta\tau}}\right)\right] = \int_{x(0)=x_i}^{x(\beta)=x_f}\mathcal{D}x\,, \quad \text{(I.21)}$$

with $\beta=(N+1)\Delta\tau$. Thus the configuration-space PI representation of the density matrix (or imaginary-time PI) corresponds to a "sum" over all continuous trajectories $x(\tau)$, $\tau\in[0,\beta]$, connecting the initial point $x(0)=x_i$ with the final point $x(\beta)=x_f$.

In connection with (I.20) it should be noted that the integral $\int_0^\beta d\tau \left[\frac{m}{2}\dot{x}^2 + V(x)\right]$ is the classical *Euclidean action* integral[6] along the path $x(\tau)$ with $\tau \in [0, \beta]$. One can also regard (I.20) as an expectation value of the functional $\exp[-\int_0^\beta d\tau\, V(x(\tau))]$ over the (driftless) Brownian motion with the diffusion coefficient $1/m$, and duration β, that starts at point x_i, and terminates at x_f. Corresponding stochastic process with fixed end points is also known as *Brownian bridge*.

Equation (I.20) is much better behaved mathematically than (I.13). In particular, the new exponential tends to suppress very rough paths. Even so, the "typical" path $x(\tau)$ is merely continuous and not differentiable. This is because with the time sliced density and (I.16) the mean square velocity diverges when the time slicing $\Delta\tau \to 0$, namely

$$\langle v^2 \rangle \;\equiv\; \left\langle \left(\frac{x_{k+1} - x_k}{\Delta\tau} \right)^2 \right\rangle \;=\; \frac{1}{\beta m \Delta\tau}\,. \tag{I.22}$$

This means that a typical path in the PI is a nowhere differentiable continuous path — fractal, with the *Hurst exponent* $H = 1/2$ and the *Hausdorff–Besicovich dimension* $D_H = 2 - H = 3/2$. So, $\dot{x}^2$ appearing in the exponent should not be taken too literally and in fact mathematicians typically subsume it into the measure, which is then known as *Wiener measure*. The latter is defined as

$$d\mu_W^m(x) \;\equiv\; \exp\left\{ -\frac{1}{2m} \int_0^\beta d\tau\, \dot{x}^2(\tau) \right\} \mathcal{D}x\,, \tag{I.23}$$

where $\mathcal{D}x$ is given in (I.21). Strictly speaking, (I.23) represents the so-called *conditional* Wiener measure, i.e., measure on a set of continuous trajectories $x(\tau)$, with $\tau \in [0, \beta]$, for which $x(0) = x_i$ and $x(\beta) = x_f$. The traditional *uncoditional* Wiener measure, which we do not really use here, specifies its values only at the initial time, and it corresponds to one additional integration (i.e., extra dx_f in $\mathcal{D}x$) over the final value x_f at the final time t_f.

Let us finally note that despite the similarity of the PI (I.20) with the Feynman PI (I.13), there is a profound difference between these two expressions. In the former (Feynman's) case the underlying measure is only *finitely additive* (the limiting procedure is ambiguous), while in the latter

[6]The adjective *Eucliden* simply emphasizes the fact that the time derivative term in the action has an opposite sign in comparison to the original time derivative term in (I.13). This terminology descents from Quantum Field Theory where one starts with Minkowski metric in the gradient term and by changing the sign in front of a time derivative term one obtains (modulo overall minus sign) an Euclidean metric in the gradient part.

(Wiener's) case the continuum limit actually defines a genuine measure, i.e., a *countably additive measure* on paths. In fact, the Wiener measure can be formulated even without time-slicing procedure [104, 135, 235, 282]. Various attempts to put Feynman's PI on a firm mathematical ground, e.g., define Feynman's PI with respect to a *pseudomeasure* [54, 176, 230] are still under way.

Appendix J

Pawula theorem

In this Appendix I will prove Pawula theorem. My exposition will loosely follow Refs. [249, 269].

In its essence, the Pawula theorem states that there are only three possible cases in the Kramers–Moyal expansion (3.40)-(3.38): (a) The Kramers–Moyal expansion is truncated at $n = 1$, which means that the process is *deterministic*; (b) the Kramers–Moyal expansion stops at $n = 2$, in which case it is equivalent to the Fokker–Planck equation, and describes diffusion processes and, finally, (c) the Kramers–Moyal expansion contains all coefficients up to $n = \infty$. The Pawula theorem is a consequence of the following lemma:

Lemma J.1. *If $D^{(n)} < \infty$ for all n and if $D^{(n)} = 0$ for some even $n > 2$, then $D^{(n)} = 0$ for all $n \geq 3$. If $D^{(2)} = 0$ then only $D^{(1)}$ can be nonzero.*

Proof.

To prove the lemma, let me consider the Cauchy–Schwarz inequality

$$\left[\int dz\, p(z) f(z) g(z)\right]^2 \;=\; \left[\int dz\, p(z) f^2(z)\right]\left[\int dz\, p(z) g^2(z)\right], \quad \text{(J.1)}$$

where both $f(z)$ and $g(z)$ are non-negative functions and $p(z)$ is some probability distribution. Now I set $p(z) = P(z, t+\tau|x, t)$, $f(z) = (z-x)^n$ and $g(z) = (z-x)^{n+m}$, with $n \geq 0$ and $m \geq -n$, and divide both sides of (J.1) by τ^2. In the limit of $\tau \to 0$ I obtain

$$\left[(2n+m)! D^{(2n+m)}\right]^2 \;\leq\; (2n!)(2n+2m)! D^{(2n)} D^{(2n+2m)}\,. \quad \text{(J.2)}$$

So, in particular

$$\begin{aligned} D^{(n)} &\propto D^{(2n)}D^{(0)}\,, \\ D^{(n+1)} &\propto D^{(2n)}D^{(2)}\,, \\ D^{(n+2)} &\propto D^{(2n)}D^{(4)}\,, \\ &\vdots \end{aligned} \tag{J.3}$$

By noting the fact that $D^{(0)}$ diverges with $1/\tau$ as $\tau \to 0$, one must exclude the first line in (J.3). Assuming that $D^{(4)} = 0$ (i.e., $n = 2$), then from (J.3) I have

$$D^{(4)} = 0 \Rightarrow D^{(3)} = D^{(5)} = D^{(6)} = \cdots = 0\,. \tag{J.4}$$

By assuming that $D^{(2)} = 0$ (i.e., $n = 1$), I similarly get

$$D^{(2)} = 0 \Rightarrow D^{(3)} = D^{(5)} = D^{(6)} = \cdots = 0\,. \tag{J.5}$$

Note that this does not provide any restriction on $D^{(1)}$, so when $D^{(2)} = 0$ then only $D^{(1)}$ can be nonzero. This closes the proof. □

Clearly, the only possibility how no to have all $D^{(n)}$ with $n \geq 3$ equal zero is that there should not exist any non-zero $D^{(n)}$ with n even. This, in turn means that all terms up to $n = \infty$ must be considered, though, some of the odd coefficients can be zero.

It should be stressed finally that the Pawula theorem, does not state that the expansions truncated at $n \geq 3$ cannot be used (in fact often it is used), but in such cases the ensuing transition probability possess various undesirable properties, e.g., it has inevitably negative values at least for sufficiently short time [269].

Appendix K

Feynman chessboard — technical exposition

The relativistic checkerboard (also chessboard) picture [92, 137, 165] was an attempt by R.P. Feynman to generalize his space-time approach to include *special relativity*. In particular, Feynman hoped that he could explain particle's spin as a result of the space-time structure alone. This attempt was in many respects incomplete, though some recent works suggest that the checkerboard picture may prove to be more fundamental than just an interesting mathematical curiosity [149, 159, 241, 242]. In this Appendix I provide a brief technical exposition of the basic ideas involved.

K.1 Prelude — Dirac equation in $1+1$ dimensions

In ordinary $3+1$ dimensions the Dirac equation can be written in a Schrödinger-like form [72]:

$$i\hbar\frac{\partial\psi}{\partial t} \;=\; mc^2\gamma^0\psi - i\hbar c\gamma^0\gamma^i\frac{\partial\psi}{\partial x^i} \;=\; mc^2\beta\psi - i\hbar c\alpha^i\frac{\partial\psi}{\partial x^i}\,, \qquad \text{(K.1)}$$

with the γ-matrices satisfying the Clifford algebra $\{\gamma^\mu,\gamma^\nu\} = 2\eta^{\mu\nu}$. In passing to 1+1 dimensions the Clifford algebra keeps its structure, but μ and ν can be only 0 or 1. The corresponding representation of the two γ-matrices is well known. In particular, one can choose $\gamma^0 = \gamma_0 = \sigma_1$ and $\gamma_1 = -\gamma^1 = i\sigma_2$. With this the Dirac equation in $1+1$ dimensions can be cast in the form

$$i\hbar\frac{\partial\psi}{\partial t} \;=\; mc^2\sigma_1\psi - i\hbar c\sigma_1(-i\sigma_2)\frac{\partial\psi}{\partial x} \;=\; mc^2\sigma_1\psi - i\hbar c\sigma_3\frac{\partial\psi}{\partial x}\,, \quad \text{(K.2)}$$

where I have used the identity $\sigma_i\sigma_j = \delta_{ij} + i\varepsilon_{ijk}\sigma_k$.

The propagator, say $G(x,t_x;y,t_y)$, for ψ is a 2×2 matrix function of space and time which fulfills the defining equation

$$\psi(x,t_x) \;=\; \int_{\mathbb{R}} dy\, G(x,t_x;y,t_y)\,\psi_0(y,t_y)\,. \tag{K.3}$$

Euclidean version of (K.2) is easily obtained when one analytically continues the time t to imaginary times and assumes the Clifford algebra in the form $\{\gamma_E^\mu,\gamma_E^\nu\} = 2\delta^{\mu\nu}$. By setting $t = -i\tau$ ($\tau \in \mathbb{R}$) and using $\gamma_E^0 = \sigma_1$ and $\gamma_E^1 = -\sigma_2$ one obtains

$$\frac{\partial\psi_E}{\partial\tau} \;=\; \frac{mc^2}{\hbar}\,\sigma_1\psi_E - c\sigma_3\frac{\partial\psi_E}{\partial x}\,. \tag{K.4}$$

Here $\psi_E(x,\tau) \equiv \psi(x,t=-i\tau)$. Among others, the equation (K.4) implies the correct Euclidean Klein–Gordon equation in 1+1 dimensions. The Euclidean Green function $G(x,-i\tau_x,y,-i\tau_y) \equiv P(x,\tau_x|y,\tau_y)$ satisfies the Fokker–Planck-like equation:

$$(\partial_\tau + c\sigma_3\partial_x - \sigma_1 mc^2/\hbar)\mathcal{P}(x,\tau|x',\tau') \;=\; \delta(x-x')\delta(\tau-\tau')\,. \tag{K.5}$$

Note also that $mc^2/\hbar = 1/\tau_C$ where τ_C is the Compton time, i.e., time in which light crosses a distance equal to particle's Compton wave length λ_C.

K.2 Euclidean checkerboard picture

I start by introducing a so-called *Euclidean checkerboard picture*. To this end I consider a particle with a fixed speed v moving on a line. I suppose, further, that from time to time the particle suffers a complete reversal of the direction (and hence also $v \leftrightarrow -v$). Let these reversals be random with a fixed rate, say a, of the reversal and with the probability for the reversal in a time interval $d\tau$ being $ad\tau$. Such a process represents the Poisson stochastic process.

Let $p_+(x,\tau)$ and $p_-(x,\tau)$ be PDF's for the particle being at the positions x at time τ and moving to the *right* and *left*, respectively. One can easily persuade himself that the following difference master equations hold

$$\begin{aligned} p_\pm(x,\tau+\Delta\tau) \;&=\; p_\pm(x\mp\Delta x,\tau)(1-a\Delta\tau) \\ &+\; p_\mp(x\pm\Delta x,\tau)a\Delta\tau\,. \end{aligned} \tag{K.6}$$

In the continuum limit, I receive two interlocked differential equations

$$\frac{\partial p_\pm(x,\tau)}{\partial\tau} \;=\; -a\,(p_\pm(x,\tau) - p_\mp(x,\tau)) \;\mp\; v\frac{\partial p_\pm(x,\tau)}{\partial x}\,, \tag{K.7}$$

where the identification $v = dx/d\tau$ was made. At this point one may introduce the probability doublet

$$P(x,\tau) \;=\; \begin{pmatrix} p_+(x,\tau) \\ p_-(x,\tau) \end{pmatrix}. \tag{K.8}$$

In terms of P the equation (K.7) reads as

$$\frac{\partial P(x,\tau)}{\partial \tau} \;=\; -aP(x,\tau) + a\sigma_1 P(x,\tau) - v\sigma_3 \frac{\partial P(x,\tau)}{\partial x}\,. \tag{K.9}$$

The latter implies that the underlying stochastic process is indeed the Poisson process. Equation (K.9) can be brought into a simpler form by performing the substitution

$$P(x,\tau) \;=\; e^{-a\tau}\mathcal{P}(x,\tau)\,. \tag{K.10}$$

For $\mathcal{P}$ then the following master equation holds

$$\frac{\partial \mathcal{P}(x,\tau)}{\partial \tau} \;=\; a\sigma_1 \mathcal{P}(x,\tau) - v\sigma_3 \frac{\partial \mathcal{P}(x,\tau)}{\partial x}\,. \tag{K.11}$$

Before I proceed further, two comments are in order. First, the substitution (K.10) represents more than a simple mathematical trick. In fact, back in the Minkowski picture the substitution (K.10) represents a multiplication of the wave function by a phase factor $\exp\big[(imc^2/\hbar)t\big]$. This is the usual relativistic phase factor which is responsible for a shift of the ground-state energy by mc^2 [149]. Second, by comparing Eq. (K.4) with Eq. (K.11) one can see that the equations are identical provided one sets $\mathcal{P}(x,\tau) = \psi_E(x,\tau)$, $a = mc^2/\hbar = 1/\tau_C$ and $v = c$.

The *Euclidean propagator*, or equivalently, *conditional probability* associated with Eq. (K.11) can be calculated by various means. In particular, the Feynman–Kac formula allows to phrase the propagator via path integral. There one should perform a weighted sum over *all* continuous trajectories running from the initial position (x',τ') to the final position (x,τ) (see, e.g., Ref. [179] for a detailed exposition of the path-integral approach). Interestingly enough, not all trajectories have a non-trivial contribution. Indeed, in deriving (K.11) I have considered only a special sub-class of zig-zag paths. From (K.10) follows that the propagator can be then written as

$$\mathcal{P}_{ij}(x,\tau|x',\tau') \;=\; e^{aT} P_{ij}(x,\tau|x',\tau')\,, \tag{K.12}$$

where $T \equiv \tau - \tau'$ and $P(x,\tau|x',\tau')$ is the matrix-valued conditional probability density linking together the initial (prior) PDF's $P(x',\tau')$ and the final (marginal) PDF $P(x,\tau)$, i.e.

$$P_i(x,\tau) \;=\; \int_{\mathbb{R}} dx' P_{ij}(x,\tau|x',\tau')\, P_j(x',\tau')\,. \tag{K.13}$$

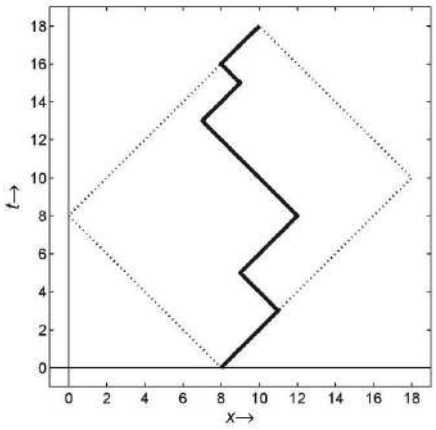


Fig. K.1: An example of a typical "++" trajectory that enters Feynman checkerboard picture for a spin 1/2 particle in (1+1)-dimensional spacetime.

Indices i, j in both (K.12) and (K.13) run through the set $\{+, -\}$. At this stage I define X_{ij} as a set of all Poisson processes (with the rate a) that start at the state described by values $\{x', \tau', j\}$ and end at the state $\{x, \tau, i\}$. I denote $n(T)$ as the number of events that happen in the time interval T. With these I can write the transitional probability for our Poisson process by summing (marginalizing) in the the joint probability distribution $P((n(T) = k) \cap X_{ij})$ over all possible events k. In particular

$$\begin{aligned} P_{ij}(x, \tau | x', \tau') &= \sum_k P((n(T) = k) \cap X_{ij}) \\ &= \sum_k \mathfrak{P}(n(T) = k) P(X_{ij} | n(T) = k)\,. \end{aligned} \tag{K.14}$$

The second equality in (K.14) is just the Bayes rule with $\mathfrak{P}(n(T) = k)$ describing the probability that during the time T one observes k events. The analysis can be carried further when I discretize the time interval into N equidistant pieces. Such a *regularization* allows to evaluate the above conditional probability explicitly for any N. As with any regulating scheme, the limit $N \to \infty$ should be performed *after* the calculations with the fixed

N are done. Consequently I can write

$$P_{ij}(x,\tau|x',\tau') \;=\; \lim_{N\to\infty}\sum_k \mathfrak{P}(n(T)=k)\,\frac{\phi^{(N)}((n(T)=k)\cap X_{ij})}{\phi^{(N)}(n(T)=k)}$$

$$=\; \lim_{N\to\infty} e^{-aT}\sum_k \left(\frac{aT}{N}\right)^k \phi^{(N)}((n(T)=k)\cap X_{ij})\,. \quad \text{(K.15)}$$

In the first equality I have utilized the definition of the conditional probability. The function $\phi^{(N)}(\ldots)$ denotes the *number* of distinct realizations of the random process in its argument. In the second equality I have used the fact that in the large N limit Stirling's formula gives

$$\phi^{(N)}(n(T)=k) \;=\; \frac{N!}{k!(N-k)!} \;\cong\; \frac{N^k}{k!}\,. \quad \text{(K.16)}$$

This in turn allows to write the Euclidean propagator in the form

$$\mathcal{P}_{ij}(x,\tau|x',\tau') \;=\; \lim_{N\to\infty}\sum_k \left(\frac{mc^2T}{N\hbar}\right)^k \phi_{ij}^{(N)}(k)\,. \quad \text{(K.17)}$$

Here I have employed the simplifying notation $\phi_{ij}^{(N)}(k) \equiv \phi^{(N)}((n(T)=k)\cap X_{ij})$. Eq. (K.17) is (in spirit of Feynman's PI) a sum over alternatives, where each summand is a (weighted) number of the Poisson processes (with the rate $a = mc^2/\hbar$) with exactly k reversals (events) running from the initial state $\{x',\tau',j\}$ to the final state $\{x,\tau,i\}$.

Since $\phi_{ij}^{(N)}(k)$ represents the number of paths it does not change when I pass from Euclidean to Minkowski picture. The only change is in the path weighting factor $(mc^2T/N\hbar)^k$ where the Euclidean T changes to the Minkowski iT. Consequently I can write the ensuing Minkowski Green function in the form

$$G_{ij}(x,t|x',t') \;=\; \lim_{N\to\infty}\sum_k \left(i\frac{mc^2T}{N\hbar}\right)^k \phi_{ij}^{(N)}(k)\,, \quad \text{(K.18)}$$

where now $T = t - t'$. This is precisely Feynman's original result [92]. By its very construction should (K.18) coincide with the usual Bessel-functions based formula for the Dirac propagator. Taking into account the explicit form of $\phi_{ij}^{(N)}(k)$ this can be, indeed, easily checked. Following Refs. [137, 165], I can calculate, for instance, $\phi_{++}^{(N)}(k)$ as follows: first I realize that for any "++" (and also "−−") configuration k is an even number and there are precisely $k/2$ turns to the left and $k/2$ turns to the right.[7] By

[7] For "−+" and "+−" configurations k is odd. For "−+" configuration there $(k-1)/2+1$ turns to the left and $(k-1)/2$ turns to the right. For "+−" configuration the rôle of *left* and *right* is reversed.

splitting the time axis into N equidistant pieces, I define the elementary time step $\Delta\tau \equiv T/N$. I further denote the total time (in the units of $\Delta\tau$) that corresponds to the left turns as L and the time for right turns as R. With this I can write

$$x - x' \;=\; c(R-L)\Delta\tau \;\equiv\; c\Delta\tau M \quad \text{and} \quad R+L \;=\; N\,, \tag{K.19}$$

$$\Rightarrow \quad R \;=\; \frac{N+M}{2}\,, \quad L \;=\; \frac{N-M}{2}\,. \tag{K.20}$$

From Fig. K.2 one may infer that the total number of "++" paths starting at (x', τ') and ending at (x, τ) with k turns can be written as the product

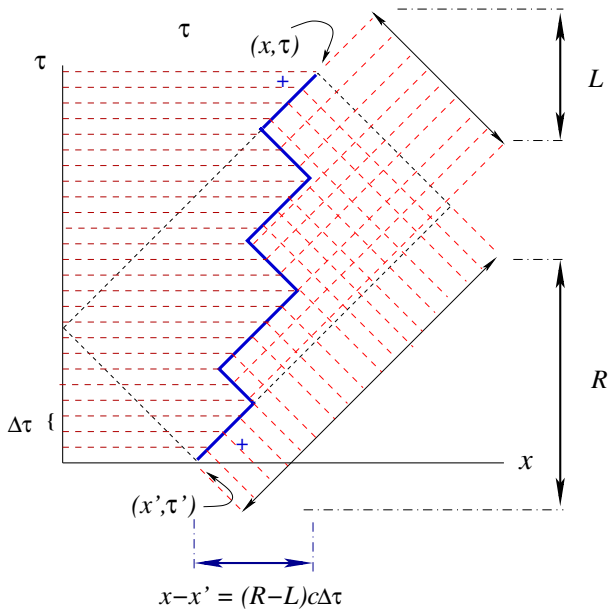


Fig. K.2: Conventions used in the $\phi^{(N)}_{++}(k)$ calculation.

---

of the number of different ways in which one can distribute $k/2$ right turns among $R-2$ potential time occasions for the right turn (-2 enters because the initial and final time steps are fixed by the "++" boundary conditions), times the number in which one can distribute $k/2-1$ left turns (-1 enters because one turn to the left is compulsory for $m \neq 0$) among $L-1$ potential time occasions for the left turn (-1 enters because one turn to the left is

compulsory for $m \neq 0$), i.e.

$$\phi^{(N)}_{++}(k) \;=\; \binom{R-2}{k/2}\binom{L-1}{k/2-1} \;\cong\; \frac{L^{k/2-1}}{(k/2-1)!}\frac{R^{k/2}}{(k/2)!}$$

$$=\; \frac{(LR)^{(k-1)/2}}{(k/2-1)!(k/2)!}\sqrt{\frac{R}{L}}\,. \tag{K.21}$$

Using further that

$$LR \;=\; \frac{1}{4}\left(N^2-M^2\right) \;=\; \frac{1}{4}N^2\left(1-v^2/c^2\right) \;=\; \left(\frac{N}{2\gamma}\right)^2, \tag{K.22}$$

$$R/L \;=\; \frac{(1+M/N)}{(1-M/N)} \;=\; \frac{(1+v/c)}{(1-v/c)}$$

$$=\; \frac{\gamma^2}{(\Delta\tau^2 N)^2}\left(T+(x-x')/c\right)^2, \tag{K.23}$$

(here $v=(x-x')/T$ and γ is the Lorentz factor) and plugging (K.22)–(K.23) to (K.21) I receive

$$\mathcal{P}_{++}(x,t|x',t') \;\cong\; \frac{\gamma}{T}\left(T+(x-x')/c\right)\sum_{k=1}^{\infty}\left(\frac{mc^2T}{N\hbar}\right)^{2k}\left(\frac{N}{2\gamma}\right)^{2k-1}\frac{1}{(k-1)!k!}$$

$$=\; \Delta\tau\,\frac{mc^2(T+(x-x')/c)}{s\hbar}\,I_1(mc^2s/\hbar)\,. \tag{K.24}$$

In the last line I have used the series representation for the Modified Bessel function [1]

$$I_1(z) \;=\; \frac{z}{2}\sum_{n=0}^{\infty}\left(\frac{z}{2}\right)^{2n}\frac{1}{n!(n+1)!}\,, \tag{K.25}$$

and abbreviation $s=T/\gamma=(T^2-(x-x')^2/c^2)^{1/2}$. In a similar way I would arrive at the result

$$\mathcal{P}_{-+}(x,t|x',t') \cong\; \Delta\tau\,\frac{mc^2}{\hbar}\,I_0(mc^2s/\hbar)\,. \tag{K.26}$$

The correct normalization in the continuum limit can be deduced from the probability normalization. Going back to the Poisson process described by $P_{ij}(x,\tau|x',\tau')$ (cf. Eq. (K.12)) one should require the normalization

$$e^{-at}\int_{-ct}^{ct}dx\;\left[\mathcal{P}_{-+}(x,t|0,0)+\mathcal{P}_{++}(x,t|0,0)\right] \;=\; P_+(n(t)\geq 1)$$

$$=\; 1-\exp(-at)\,, \tag{K.27}$$

(similar normalization should hold if I have started from the "−" initial configuration). Here I have used the fact that the LHS of the equation represents the probability that the particle (which starts from the state $\{0, 0, +\}$) is in the interval $(-ct, ct)$ at time t. At the same time this should represent probability that during t will happen at least one reversal. For the Poisson process in question this is $1 - e^{-at}$. Explicit calculation, however, gives that

$$e^{-at}\int_{-ct}^{ct} dx\ [\mathcal{P}_{-+}(x,t|0,0) + \mathcal{P}_{++}(x,t|0,0)] \ = \ 2c\Delta\tau(1-e^{-at})\,. \quad \text{(K.28)}$$

This suggests that both (K.24) and (K.26) should be divided by $2c\Delta\tau$ in order to fulfill the normalization condition (K.27).

The passage from the Euclidean transition probability (K.24) to the Minkowski Green function is established by changing T to iT in the path-weighting factor. Consequently I can write

$$G_{++}(x,t|x',t') \ \cong \ -\frac{mc(T+(x-x')/c)}{2s\hbar}\, J_1(mc^2 s/\hbar)\,, \qquad \text{(K.29)}$$

where I have employed the series representation for the Bessel function of the first kind [1]

$$J_1(z) \ = \ \frac{z}{2}\sum_{n=0}^{\infty}(-1)^n \left(\frac{z}{2}\right)^{2n} \frac{1}{n!(n+1)!}\,, \qquad \text{(K.30)}$$

Analogous reasonings can be done also for other components. Finally I can write for the full matrix propagator

$$G_{ij}(x,t|x',t') \ =$$

$$\frac{mc}{2s\hbar}\begin{pmatrix} -(T+(x-x')/c)J_1(mc^2 s/\hbar) & s\,J_0(mc^2 s/\hbar) \\ s\,J_0(mc^2 s/\hbar) & (-T+(x-x')/c)J_1(mc^2 s/\hbar) \end{pmatrix}. \qquad \text{(K.31)}$$

This indeed coincides with the known result for the 1+1 dimensional Dirac's fermion propagator (see, e.g. Ref. [137]).

Eq. (K.17) offers a conceptually interesting interpretation of a representative trajectory for an Euclideanized Dirac particle in two dimensions. In particular, a massive(!) particle propagates over an average distance λ_C with velocity c before it reverts its direction. It is only on much larger spatial scales (after many directional reversals take place) where the Brownian motion with a sub-luminal average velocity emerges. So in quantum mechanics one must sum over all such zig-zag trajectories subject given

Dirichlet boundary conditions in order to obtain correct Feynman causal propagator for Dirac's fermion.

The results obtained should be compared with the corresponding Euclidean version for a non-relativistic particle. In this case the particles obey Wiener (i.e., Brownian) process, where the scaling $\langle \Delta x^2 \rangle \propto \Delta t$ implies that the representative trajectories have the Hurst exponent $H = 1/2$, or equivalently the Hausdorff fractal dimension D_H of the trajectory is $3/2$. In addition, such a scaling is independent of the local potential as long as the potential is velocity independent. In other words, all local potentials fall into the same universality class as a free particle [186]. This is also reflected in the fact that the canonical commutation relations for non-relativistic particles have all the same algebraic structure.

Bibliography

[1] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions* (Dover Publications, New York).

[2] Aczél, J. and Darótzy, Z. (1975). *Measure of Information and Their Characterizations* (Academic Press, New York).

[3] Ahlswede, R. and Cai, N. (2006). An interpretaion of identification entropy, *IEEE Trans. Inf. Theory* **52**, p. 4198.

[4] Albaverio, S. A., Høegh-Krohn, R. and Mazzucchi, S. (2008). *Mathematical Theory of Feynman Path Integrals: An Introduction* (Springer, Berlin).

[5] Amari, S. (2016). *Information Geometry and Its Applications* (Springer, London, U.K.).

[6] Ambjorn, J., Durhuus, B. and Jonsson, T. (2005). *Quantum Geometry: A Statistical Field Theory Approach* (Cambridge University Press, Cambridge, U.K.).

[7] Ambjorn, J., Jurkiewicz, J. and Loll, R. (2004). Emergence of a 4D world from causal quantum gravity, *Phys. Rev. Lett.* **93**, p. 131301.

[8] Ash, R. (1965). *Information Theory* (Wiley, New York).

[9] Ashcroft, N. and Mermin, N. (1976). *Solid State Physics* (Saunders College, Philadelphia).

[10] Babenko, K. I. (1962). An inequality in the theory of Fourier integrals, *Amer. Math. Soc. Transl.* **44**, p. 115.

[11] Badii, R. and Politi, A. (1985). Statistical description of chaotic attractors, *J. Stat. Phys.* **40**, p. 725.

[12] Bak, P., Tang, C. and Wiesenfeld, K. (1988). Self-organized criticality, *Phys. Rev.* **A38**, p. 364.

[13] Balescu, R. (1997). *Statistical Dynamics: Matter out of equilibrium* (Imperial College Press, London).

[14] Ball, R., Nauenberg, M. and Witten, T. A. (1984). Diffusion-controlled aggregation in the continuum approximation, *Phys. Rev.* **A29**, p. 2017.

[15] Bar-Hillel, Y. and Carnap, R. (1953). Semantic information, *British J. Philos. Sci.* **14**, p. 146.

[16] Barndorff-Nielsen, O. (1963). On the limit of extremee order statistics, *Annals of Math. Stat.* **34**, p. 992.

[17] Barnett, L., Barrett, A. B. and Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables, *Phys. Rev. Lett.* **103**, p. 238701.
[18] Barron, A. R. (1986). Entropy and the central limit theorem, *Ann. Probab.* **14**, p. 336.
[19] Bashkirov, A. G. and Sukhanov, A. D. (2002). The distribution function for a subsystem experiencing temperature fluctuations, *J. Exp. Theor. Phys.* **95**, p. 440.
[20] Bastianelli, F. and van Nieuwenhuizen, P. (2009). *Path Integrals and Anomalies in Curved Space* (Cambridge University Press, Cambridge).
[21] Beck, C. (2004a). Generalized statistical mechanics of cosmic rays, *Physica A* **331**, p. 173.
[22] Beck, C. (2004b). Superstatistics in hydrodynamic turbulence, *Physica D: Nonlinear Phenomena* **193**, p. 195.
[23] Beck, C. (2007). Statistics of three-dimensional Lagrangian turbulence, *Phys. Rev. Lett.* **98**, p. 064502.
[24] Beck, C. and Cohen, E. G. D. (2003). Superstatistics, *Physica A*. **322**, p. 267.
[25] Beck, C., Cohen, E. G. D. and Swinney, H. L. (2005). From time series to superstatistics, *Phys. Rev. E* **72**, p. 056133.
[26] Beck, C. and Schloegl, F. (1993). *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, U.K.).
[27] Beckner, W. (1975). Inequalities in fourier analysis, *Annals of Mathematics* **102**, p. 159.
[28] Bengtsson, I. and Życzkowski, K. (2006). *Geometry of Quantum States. An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, UK).
[29] Bennett, C. H. (1982). The thermodynamics of computation - a review, *Int. J. Theor. Phys.* **21**, p. 905.
[30] Bennett, C. H. (1987). Demons, engines and the second law, *Scientific American* **11**, p. 108.
[31] Bennett, C. H., Brassard, G., Crépeau, C. and Maurer, U. M. (1995). Generalized privacy amplification, *IEEE Transactions on Information Theory* **41**, p. 1915.
[32] Bercheri, J. F. (2009). Source coding with escort distributions and Rényi entropy bounds, *Phys. Lett. A* **373**, p. 3235.
[33] Bernoulli, D. (1738 and 2004). *Hydrodynamica* (Dover Publications, London).
[34] Bertoin, J. (1996). *Lévy Processes* (Cambridge University Press, Cambridge).
[35] Białynicki-Birula, I. (1985). Entropic uncertainty relations in quantum mechanics, in *Quantum Probability and Applications II* eds. L. Accardi and W. von Waldenfels, Lecture Notes in Mathematics 1136 (Springer, Berlin).
[36] Białynicki-Birula, I. and Mycielski, J. (1975). Uncertainty relations for information entropy in wave mechanics, *Commun. Math. Phys.* **44**, p. 129.
[37] Billingsley, P. (1965). *Ergodic Theory and Information* (Wiley, New York).

[38] Birkhoff, G. D. (1931). Proof of the ergodic theorem, *Proc. Nat. Acad. Sci. USA* **17**, p. 656.
[39] Blasone, M., Jizba, P. and Kleinert, H. (2005a). Path integral approach to 't Hooft's derivation of quantum from classical physics, *Phys. Rev.* **A71**, p. 052507.
[40] Blasone, M., Jizba, P. and Kleinert, H. (2005b). Quantum behavior of deterministic systems with information loss. Path integral approach, *Ann. Phys.* **320**, p. 468.
[41] Blasone, M., Jizba, P. and Vitiello, G. (2011). *Quantum Field Theory And Its Macroscopic Manifestations* (Imperila College Press, London).
[42] Boettcher, F., Renner, C., Waldl, H. P. and Peinke, J. (2003). On the statistics of wind gusts, *Boundary-Layer Meteorology* **108**, p. 163.
[43] Bogolyubov, N. N. and Krylov, N. M. (1937). La théorie générate de la mesure dans son application a l'étude de systèmes dynamiques de la mécanique non-linéaire. *Ann. Math. II* **38**, p. 65.
[44] Boltzmann, L. (1877). Ueber die Beziehung dem zweiten Haubtsatze der mechanischen Waermetheorie und der Wahrscheinlichkeitsrechnung respektive den Saetzen ueber das Waermegleichgewich, *Wiener Berichte* **76**, p. 373.
[45] Bouchaud, J.-P. and Potters, M. (2001). *Theory of financial risks, from statistical physics to risk management* (CUP, Cambridge).
[46] Brillouin, L. (1956). *Science and Information Theory* (Academic, New York).
[47] Buchdahl, H. A. (1958). A formal treatment of the consequences of the second law of thermodynamics in Carathéodory's formulation, *Zeit. Phys.* **152**, p. 425.
[48] Burbeal, J. (1986). Informative geometry of probability spaces, *Expo. Math.* **4**, p. 347.
[49] Burbeal, J. and Rao, C. (1982). Differential metrics in probability spaces, *Probability Math. Statist.* **3**, p. 115.
[50] Campbell, L. L. (1965). A coding theorem and Rényi's entropy, *Information and Control* **8**, p. 423.
[51] Campbell, L. L. (2000). A block coding theorem and Rényi's entropy, *IEEE Transactions on Information Theory* **41**, p. 338.
[52] Caratheodory, C. (1909). Untersuchungen über die Grundlagen der Thermodynamik, *Math. Ann.* **67**, p. 355.
[53] Carruthers, P. and Nietto, M. M. (1968). Phase and angle variables in quantum mechanics, *Rev. Mod. Phys.* **40**, p. 411.
[54] Cartier, P. and DeWitt-Morette, C. e. (2006). *Functional Integration: Action and Symmetries* (Cambridge University Press, BCambridge).
[55] Celeghini, E., De Martino, S., De Siena, S., Rasetti, M. and Vitiello, G. (1995). Quantum groups, coherent states, squeezing and lattice quantum mechanics, *Ann. Phys.* **241**, p. 50.
[56] Chaitin, G. (1987). *Algorithmic Information Theory* (Cambridge University Press, Cambridge, U.K.).
[57] Chandrasekhar, S. (1957). *Stellar Structure* (Dover Publications, New

York).
[58] Chen., L. L. and Beck., C. (2008). A superstatistical model of metastasis and cancer survival, *Physica A* **783**, p. 3162.
[59] Clausius, R. (1865). Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie, *Annalen der Physik* **125**, p. 353.
[60] Collaboration, P. (2016). Planck 2015 results. xiii. cosmological parameters, *Astronomy and Astrophysics* **594**, p. A13.
[61] Cones, A. (1974). *Noncommutative geometry* (Academic Press, New York).
[62] Cover, M. and Thomas, J. A. (2006). *Elements of Information Theory* (John Wiley & Sons, Inc., New York).
[63] Csiszár, I. (1995). Generalized cutoff rates and Rényi's information measures, *IEEE Trans. Inform. Theory* **26**, p. 34.
[64] Csiszár, I. and Shields, P. (2004). *Information and Statistics: A Tutorial* (Publishers Inc., Boston).
[65] Davies, P. (1993). *The Mind of God* (Simon & Schuster, London).
[66] de Finetti, B. (2006). *Selected Works Voll. I, II* (Edizioni Cremonese, Roma).
[67] De Palma, G. (2019). The entropy power inequality with quantum conditioning, *J. Phys. A Math. Theor.* **52**, p. 08LT03.
[68] Deuschel, J.-P., Stroock, D. W. and Zessin, H. (1991). Microcanonical distribution for lattice gases, *Commun. Math. Phys* **139**, p. 83.
[69] Deutsch, D. L. (1983). Uncertainty in quantum measurements, *Phys. Rev. Lett.* **50**, p. 631.
[70] DeWitt, B. (2003). *The Global Approach to Quantum Field Theory: 2-Volume Set* (Oxford University Press, Oxford, U.K.).
[71] Dirac, P. (1958). *The principles of Quantum Mechanics* (Claredon Pres, Oxford).
[72] Dirac, P. (1966). *Lectures on Quantum Field Theory* (Belfer Graduate School of Science, Monograph series, n.3, Yeshiva Univ., New York).
[73] Dirac, P. A. M. (1933). The lagrangian in quantum mechanics, *Physikalische Zeitschrift der Sowjetunion* **3**, p. 64.
[74] Dobrushin, R. L. and Tirozzi, B. (1977). The central limit theorem and the problem of equivalence of ensembles, *Commun. Math. Phys.* **54**, p. 173.
[75] Dodonov, V., Kurmyshev, E. and Man'ko, V. (1980). Generalized uncertainty relation and correlated coherent states, *Phys. Lett. A* **79**, p. 150.
[76] Dowker, J. S. (1976). Path integrals and ordering rules, *J. Math. Phys.* **17**, p. 1873.
[77] Ekert, A. and Knight, P. L. (1995). Entangled quantum systems and the schmidt decomposition, *Am. J. Phys.* **63**, p. 415.
[78] Ellis, R. (2006). *Entropy, Large Deviations, and Statistical Meclianics* (Springer, Berlin).
[79] Erhart, J., Sponar, S., Sulyok, G., Badurek, G., Ozawa, M. and Hasegawa, Y. (2012). Goldstone theorem, Hugenholtz–Pines theorem and Ward–Takahashi relation in finite volume Bose–Einstein condensed gases, *Nature Phys.* **8**, p. 185.

[80] Falkoner, K. (2003). *Fractal Geometry: Mathematical Foundations and Applications.* (Wiley, New Jersey).
[81] Feder, J. (1988). *Fractals* (Plenum Press, New York).
[82] Feinstein, A. (1958). *Foundations of Information Theory* (McGraw-Hill, New York).
[83] Feller, W. (1935). Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Math. Z.* **40**, p. 521.
[84] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications, Vol. 2* (Wiley, New York).
[85] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol. 1* (Wiley, New York).
[86] Feshbach, H. and Villars, F. (1958). Elementary relativistic wave mechanics of spin 0 and spin 1/2 particles, *Rev. Mod. Phys.* **30**, p. 24.
[87] Feynman, R. P. (1947). Space-time approach to non-relativistic quantum mechanics, *Rev. Mod. Phys.* **20**, p. 367.
[88] Feynman, R. P. (1951). Mathematical formulation of the quantum theory of electromagnetic interaction, *Phys. Rev.* **84**, p. 108.
[89] Feynman, R. P. (1966). The development of the space-time view of quantum electrodynamics, *Science* **153**, p. 699.
[90] Feynman, R. P. (1998). *Quantum Electrodynamics* (Westview Press., New York).
[91] Feynman, R. P. (2005). *Feynman's Thesis — A New Approach To Quantum Theory* (World Scientific, New York), ed. Brown, L. M.
[92] Feynman, R. P. and Hibbs, A. R. (1965). *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York).
[93] Frieden, B. R. (200). *Physics from Fisher Information* (Cambridge University Press, Cambridge, U.K.).
[94] Furstenberg, H. (1973). The unique ergodicity of the horocycle flow, in Recent advances in topological dynamics, *Springer Lecture Notes 3* **318**, p. 95.
[95] Gardner, R. J. (2002). On the symmetry function of a convex set, *Bull. Amer. Math. Soc.* **39**, p. 355.
[96] Genquoteötze, W. (2008). *Dynamics of Glass Forming Liquids* (Oxford Univ. Press, Oxford).
[97] Georgii, H. O. (1993). Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$, *Ann. Prob.* **21**, p. 1845.
[98] Ghosh, S. (2007). DSR relativistic particle in a lagrangian formulation and non-commutative spacetime: A gauge independent analysis, *Phys. Lett. B* **648**, p. 262.
[99] Giannopoulos, A. A. and Milman, V. (2000). Concentration property on probability spaces, *Advances in Mathematics* **156**, p. 77.
[100] Gibbon, E. (2001). *The History of the Decline and Fall of the Roman Empire* (Penguin Classics, London).
[101] Gibbs, J. W. (1902). *Elementary principles in statistical mechanics* (Charles Scribner's sons, New York).

[102] Girelli, F., Livine, E. R. and Oriti, D. (2005). Deformed special relativity as an effective flat limit of quantum gravity, *Nucl. Phys. B* **708**, p. 411.
[103] Girelli, F., Livine, E. R. and Oriti, D. (2010). Four-dimensional deformed special relativity from group field theories, *Phys. Rev. D* **81**, p. 024015.
[104] Glimm, J. and Jaffe, A. M. (1987). *Quantum Physics. A Functional Integral Point of View* (Springer Verlag, New York, USA).
[105] Gnedenko, B. (1940). On the theory of domains of attraction of stable laws, *Uchenye Zapiski Moskov. Gos. Univ. Matematiki* **45**, p. 61.
[106] Gnedenko, B. and Kolmogorov, A. (1954). *Limit Distributions for Sum of Random Variables* (Addison-Wesley, Reading).
[107] Gozzi, E. (1988). Hidden BRS invariance in classical mechanics, *Phys. Lett.* **B201**, p. 525.
[108] Gozzi, E., Reuter, M. and Thacker, W. D. (1989). Hidden BRS invariance in classical mechanics. ii, *Phys. Rev.* **D40**, p. 3363.
[109] Granger, C. W. J. (1969). Investigating causal relations by econometric and cross-spectral methods, *Econometrica* **37**, p. 424.
[110] Grassberger, P. (1983). Generalized dimensions of strange attractors, *Physics Letters* **A 97**, p. 227.
[111] Grassberger, P. and Procaccia, I. (1983a). Characterization of strange attractors, *Phys. Rev. Lett.* **50**, p. 346.
[112] Grassberger, P. and Procaccia, I. (1983b). Measuring the strangeness of strange attractors, *Physica* **D 9**, p. 198.
[113] Grosche, C. and Steiner, F. (1998). *Handbook of Feynman Path Integrals* (Springer, Berlin).
[114] Guias, S. (1997). *Information Theory with Applications* (McGraw-Hill, New York).
[115] Halmos, P. R. (1982). *A Hilbert Space Problem Book* (Springer-Verlag, Berlin).
[116] Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I. and Schraiman, B. I. (1986). Fractal measures and their singularities: The characterization of strange sets, *Phys. Rev.* **A33**, p. 1141.
[117] Hanel, R., Thurner, S. and Gell-Mann, M. (2011). Generalized entropies and the transformation group of superstatistics, *Proceedings of the National Academy of Sciences* **108**, p. 6390.
[118] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1934). *Inequalities* (Cambridge University Press, Cambridge).
[119] Hausdorff, F. (1919). Dimension und äußeres maß, *Math. Ann.* **79**, p. 157.
[120] Havlin, S. and Ben-Avraham, D. (1982). Diffusion on percolation clusters at criticality, *J. Phys. A: Math. Gen.* **15**, p. L691.
[121] Havlin, S. and Ben-Avraham, D. (1987). Diffusion in disordered media, *Adv. Phys* **36**, p. 695.
[122] Havrda, J. H. and Charvát, F. (1967). Quantification method of classification processes. Concept of structural α-entropy, *Kybernetica* **3**, p. 30.
[123] Hawking, S. W. (1978). *Euclidean Quantum Gravity* (Plenum, New York).
[124] Heisenberg, W. (1927). über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik, *Z. Phys.* **43**, p. 172.

[125] Hentschel, H. G. E. and Procaccia, J. (1983). The infinite number of generalized dimensions of fractals and strange attractors, *Physica D* **8**, p. 435.
[126] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Rev. Financ. Stud.* **6**, pp. 327–343.
[127] Hilhorst, H. J. (2009). Central limit theorems for correlated variables: some critical remarks, *Braz. J. Phys.* **39**, p. 371.
[128] Hirschman, I. I. (1957). A note on entropy, *Am. J. Math.* **79**, p. 152.
[129] Hlaváčková-Schindlera, K., Paluš, M., Vejmelka, M. and Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis, *Physics Reports* **441**, p. 1.
[130] Holley, R. and Stroock, D. W. (1979). Central limit phenomena of various interacting systems, *Ann. Math.* **110**, p. 333.
[131] Huang, K. (1987). *Statistical Mechanics, 2nd ed.* (John Wiley & Sons, New York).
[132] Huffman, D. (1952). A method for the construction of minimum-redundancy codes, *Proceedings of the IRE.* **40**, p. 1098.
[133] Humbert, P. (1945). Nouvelles correspondances symboliques, *Bull. Soc. Math. France* **69**, p. 121.
[134] Isham, C. I. (1995). *Lectures on Quantum Theory: Mathematical and Structural Foundations* (Imperial College Press, London).
[135] Itô, K. (1961). *Wiener integral and Feynman integral. Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability.* (California Univ. Press, Berkeley).
[136] Jackson, F. H. (1910). On q-definite integrals, *Quart.J.Pure Appl.Math.* **41**, p. 193.
[137] Jacobson, T. and Schulman, L. S. (1984). Quantum stochastics: the passage from a relativistic to a non-relativistic path integral, *J. Phys. A: Math. Gen.* **17**, p. 375.
[138] Jaynes, E. T. (1957a). Information theory and statistical mechanics, *Phys. Rev.* **106**, p. 171.
[139] Jaynes, E. T. (1957b). Information theory and statistical mechanics. II, *Phys. Rev.* **108**, p. 620.
[140] Jaynes, E. T. (1982). On the rationale of maximum-entropy methods, *Proc. IEEE.* **70**, p. 939.
[141] Jaynes, E. T. (1983). *Papers on Probability and Statistics and Statistical Physics* (D. Reidel Publishing Company, Boston).
[142] Jizba, P. (2015). Feynman checkerboard picture and neutrino oscillations, *J. Phys.: Conf. Ser.* **626**, p. 012048.
[143] Jizba, P. and Arimitsu, T. (2004a). Generalized statistics: yet another generalization, *Physica A.* **340**, p. 110.
[144] Jizba, P. and Arimitsu, T. (2004b). The world according to Rényi: Thermodynamics of multifractal systems, *Ann. Phys.* **312**, p. 17.
[145] Jizba, P. and Arimitsu, T. (2006). Towards information theory for q-nonextensive statistics without q-deformed distributions, *Ann. Phys.* **365**, p. 76.

[146] Jizba, P., Dunningham, J. A. and Joo, J. (2015). On the role of information theoretic uncertainty relations in quantum theory, *Annals of Physics* **355**, p. 2491.
[147] Jizba, P., Dunningham, J. A. and Prokš, P. (2021a). From Rényi entropy power to information scan of quantum states, *Entropy* **23**, p. 334.
[148] Jizba, P. and Kleinert, H. (2008). Superpositions of probability distributions, *Phys. Rev. E* **78**, p. 031122.
[149] Jizba, P. and Kleinert, H. (2010). Superstatistics approach to path integral for a relativistic particle, *Phys. Rev.* **D82**, p. 085016.
[150] Jizba, P., Kleinert, H. and Haener, P. (2009). Perturbation expansion for option pricing with stochastic volatility, *Physica A* **388**, p. 3503.
[151] Jizba, P., Kleinert, H. and Shefaat, M. (2012). Rényi's information transfer between financial time series, *Physica A* **391**, p. 2971.
[152] Jizba, P. and Korbel, J. (2017). On the uniqueness theorem for pseudo-additive entropies, *Entropy* **19**, p. 605.
[153] Jizba, P. and Korbel, J. (2019). Maximum entropy principle in statistical inference: Case for non-shannonian entropies, *Phys. Rev. Lett.* **122**, p. 120601.
[154] Jizba, P. and Korbel, J. (2020). When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics, *Phys. Rev.* **E101**, p. 042126.
[155] Jizba, P., Korbel, J., Lavička, H., Prokš, M., Svoboda, V. and Beck, C. (2018). Transitions between superstatistical regimes: validity, breakdown and applications, *Physica A* **493**, p. 29.
[156] Jizba, P., Korbel, J. and Zatloukal, V. (2017). Tsallis thermostatics as a statistical physics of random chains, *Phys. Rev.* **E95**, p. 022103.
[157] Jizba, P., Lavička, H. and Tabachová, Z. (2021b). Rényi transfer entropy estimators for financial time series, *Eng. Proc.* **5**, p. 33.
[158] Jizba, P., Ma, Y. and Dunningham, J. A. (2016). One-parameter class of uncertainty relations based on entropy power, *Phys. Rev.* **E93**, pp. 060104–1(R).
[159] Jizba, P. and Scardigli, F. (2012). Emergence of special and doubly special relativity, *Phys. Rev. D* **86**, p. 025029.
[160] Jizba, P. and Scardigli, F. (2013). Special relativity induced by granular space, *Eur. Phys. J. C* **73**, p. 2491.
[161] Kac, V. and Cheung, P. (2001). *Quantum Calculus* (Springer, Amsterdam).
[162] Kaniadakis, G. (2001). Non linear kinetics underlying generalized statistics, *Physica* **A296**, p. 405.
[163] Kaniadakis, G. (2005). Statistical mechanics in the context of special relativity. ii, *Phys. Rev.* **E72**, p. 036108.
[164] Kapur, J. N. (1989). *Maximum Entropy Models In Science and Engineering* (Wiley, New York).
[165] Karmanov, V. A. (1993). On the derivation of the electron propagator from a random walk, *Phys. Lett.* **A174**, p. 371.
[166] Kennard, E. H. (1927). Zur quantenmechanik einfacher bewegungtypen, *Z. Phys.* **44**, p. 326.

[167] Keylock, C. J. (2006). Constrained surrogate time series with preservation of the mean and variance structure, *Phys. Rev.* **E73**, p. 036707.
[168] Khintchine, A. Y. (1943). *Mathematical Foundations of Statistical Mechanics* (Gostekhizdat, Moscow).
[169] Khintchine, A. Y. (1957). *Mathematical Foundations of Information Theory* (Dover Publications, New York).
[170] Khintchine, A. Y. (1960). *Matematicheskie Osnovaniya Statisticheskoi Mekchaniki* (Dover Publications, London).
[171] Khintchine, A. Y. and Lévy, P. (1936). On the stable laws, *C. R. Hebd. Seances Acad. Sci* **202**, p. 374.
[172] Kimberly, D., Magueijo, J. and Medeiros, J. (2004). Nonlinear relativity in position space, *Phys. Rev. D* **70**, p. 084007.
[173] Kittel, C. (1958). *Elementary Statistical Physics* (John Wiley, New York).
[174] Kittel, C. (2004). *Introduction to Solid State Physics* (John Wiley, New York).
[175] Klafter, J., Blumen, A. and Shlesinger, M. F. (1987). Stochastic pathway to anomalous diffusion, *Phys. Rev. A* **35**, p. 3081.
[176] Klauder, J. R. (2010). *A Modern Approach to Functional Integration* (Springer, Berlin).
[177] Klauder, J. R. and Skagerstam, B. (1985). *Coherent States* (World Scientific, Singapore).
[178] Kleinert, H. (1989). *Gauge fields in condensed matter. Vol. 2: Stresses and defects. Differential geometry, crystal melting* (World Scientific, Singapore).
[179] Kleinert, H. (2009). *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World Scientific, Singapore).
[180] Kolmogorov, A. (1963). On tables of random numbers, *Sankhyā Ser.* **A 25**, p. 369.
[181] Kolmogorov, A. (1968). Logical basis for information theory and probability theory, *IEEE Transactions on Information Theory* **14**, p. 662.
[182] Kolmogorov, A. (1998). On tables of random numbers, *Theoretical Computer Science* **207**, p. 387.
[183] Kozachenko, L. F. and Leonenko, N. N. (1987). On statistical estimation of entropy of random vector, *Problems Inform. Transmission* **23**, p. 95.
[184] Kraus, K. (1987). Complementary observables and uncertainty relations, *Phys. Rev. D* **35**, p. 3070.
[185] Krein, M. and Milman, D. (1940). On extreme points of regular convex sets, *Studia Mathematica* **9**, p. 133.
[186] Kroeger, H. (2000). Fractal geometry in quantum mechanics, field theory and spin systems, *Phys. Rep.* **323**, p. 81.
[187] Kubo, R. (1988). *Statistical Mechanics, An Advanced Course with Problems and Solutions* (Nord-Holland, Amsterdam).
[188] Kubo, R., Toda, M. and Hashitsume, N. (1995). *Statistical Physics II, Nonequilibrium Statistical Mechanics* (Springer, Heidelberg).
[189] Landauer, R. (1961). Irreversibility and heat generation in the computing process, *IBM Journal of Research and Developmen* **5**, p. 183.
[190] Landbsberg, P. T. (1956). Foundations of thermodynamics, *Rev. Mod.*

Phys. **28**, p. 363.
[191] Landbsberg, P. T. (1964). A deduction of Carathéodory's principle from Kelvin's principle, *Nature* **201**, p. 285.
[192] Landbsberg, P. T. (1999). Entropies galore! *Braz. J. Phys.* **29**, p. 46.
[193] Landbsberg, P. T. and Vedral, V. (1998). Distributions and channel capacities in generalized statistical mechanics, *Phys. Lett.* **A247**, p. 211.
[194] Leonenko, N. N., Pronzato, L. and Savani, V. (2008). A class of Rényi information estimators for multidimensional densities, *The Annals of Statistics* **36**, p. 2153.
[195] Lévy, P. (1925). *Calculus des Probabilities* (Gauthier Villars, Paris).
[196] Lévy, P. (1954). *Thorie de l'addition des variables aléatoires* (Gauthier Villars, Paris).
[197] Levy-Leblond, J. M. (1976). Who is afraid of nonhermitian operators? A quantum description of angle and phase, *Ann. Phys.* **101**, p. 319.
[198] Lewis, J. T., Pfister, C. E. and Sullivan, W. G. (1994). The equivalence of ensembles for lattice systems: some examples and a counterexample, *J. Stat. Phys.* **77**, p. 397.
[199] Lieb, E. H. (1990). Gaussian kernels have only gaussian maximizers, *Invent. Math.* **102**, p. 179.
[200] Lieb, E. H. and Brascamp, H. H. (1976). Best constants in Young's inequality, its converse, and its generalization to more than three functions, *Adv. in Math.* **20**, p. 151.
[201] Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift* **15**, p. 211.
[202] Lloyd, S. (200). Quantum-mechanical Maxwell's demon, *Nature* **406**, p. 1047.
[203] Luisell, W. H. (1973). *Quantum Statistical Theory of Radiation* (Wiley, New York).
[204] Lukacs, E. (1970). *Characteristic Functions* (Hafner Publishing Co., New York).
[205] Lungerella, M., Ishigoro, K., Kuniyoshi, Y. and N. Otsu, N. (2007). Methods for quantifying the causal structure of bivariate time series, *Progress in Neurobiology* **77**, p. 1.
[206] Lyons, L. (1989). *Statistics for Nuclear and Particle Physicists* (Cambridge University Press, Cambridge).
[207] Maassen, H. and Uffink, J. B. M. (1988). Generalized entropic uncertainty relations, *Phys. Rev. Lett.* **60**, p. 1103.
[208] Mafu, M., Garapo, K. and Petruccione, F. (2013). Finite-size key in the Bennett 1992 quantum-key-distribution protocol for Rényi entropies, *Phys. Rev.* **A88**, p. 062306.
[209] Magueijo, J. and Smolin, L. (2002). Lorentz invariance with an invariant energy scale, *Phys. Rev. Lett.* **88**, p. 190403.
[210] Magueijo, J. and Smolin, L. (2003). Lorentz invariance with an invariant energy scale, *Phys. Rev. D* **67**, p. 044017.
[211] Majid, S. (2002). *A quantum groups primer* (Cambridge University Press, Cambridge).

[212] Mandelbrot, B. B. (1961). Stable paretian random functions and the multiplicative variance of income, *Econometrica* **29**, p. 517.
[213] Mandelbrot, B. B. (1963). New methods in statistical economics, *Journal of Business* **36**, p. 392.
[214] Mandelbrot, B. B. (1967). How long is the coast of britain? Statistical self-similarity and fractional dimension, *Science* **156**, p. 636.
[215] Mandelbrot, B. B. (1969). Long-run linearity, locally Gaussian processes, H-spectra and infinite variances, *Int. Econ. Rev.* **10**, p. 82.
[216] Mandelbrot, B. B. (1977). *Fractal-Form, Chance and Dimension* (Freeman, San Francisco).
[217] Mandelbrot, B. B. (1997). *Fractals and Scalling in Finance: Discontinuity, Concentration, Risk* (Springer-Verlag, New York).
[218] Mandelbrot, B. B. (1999). A fractal walk down wall street, *Scientific American* **2**, p. 70.
[219] Mandelbrot, B. B. (2001). Stochastic volatility, power-laws and long memory, *Quant. Fin.* **1**, p. 558.
[220] Mandelbrot, B. B. (2005). *Fraktale und Finanzen, Maerkte zwischen Risiko, Rendite und Ruin* (Piper, Muenchen).
[221] Mandelstam, L. and Tamm, I. G. (1945). The uncertainty relation between energy and time in non-relativistic quantum mechanics, *J. Phys. USSR* **9**, p. 249.
[222] Mantegna, R. N. and Stanley, E. (2007). *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge).
[223] Marschinski, R. and Kantz, H. (2002). Analysing the information flow between financial time series, *Eur. Phys. J.* **B30**, p. 275.
[224] Maurer, U. and Wolf, S. (1997). Privacy amplification secure against active adversaries, in B. S. Kaliski (ed.), *Advances in Cryptology — CRYPTO '97* (Springer Berlin Heidelberg, Berlin, Heidelberg), pp. 307–321.
[225] McKeon, D. G. C. and Ord, G. N. (1992). Time reversal in stochastic processes and the Dirac equation, *Phys. Rev. Lett.* **69**, p. 3.
[226] Mignemi, S. (2003a). On the definition of velocity in theories with two observer-independent scales, *Phys. Lett. A* **316**, p. 173.
[227] Mignemi, S. (2003b). Transformations of coordinates and Hamiltonian formalism in deformed special relativity, *Phys. Rev. D* **68**, p. 065029.
[228] Minsky, M. L. (1968). *Matter Mind and Models. In Minsky, M. L., ed., Semantic Information Processing* (MIT Press, Cambridge, MA.).
[229] Minsky, M. L. (2006). *The Emotion Machine* (Pantheon, New York).
[230] Morette DeWitt, C. (1972). Feynman's path integral: Definition without limiting procedure, *Commun. Math. Phys.* **28**, p. 47.
[231] Morse, P. M. and Feshbach, H. (1953). *Methods of Theoretical Physics* (McGraw-Hill Book Co., London).
[232] Müller-Lennert, M., Dupuis, F., Szehr, O. and Tomamichel, M. (2013). On quantum Rényi entropies: a new generalization and some properties, *J. Math. Phys.* **54**, p. 122203.
[233] Mustard, D. (1991). Uncertainty principles invariant under the fractional

Fourier transform, *Math. Soc. Ser. B* **33**, p. 180.
[234] Naudts, J. (2011). *Generalised Thermostatistics* (Springer, London).
[235] Nelson, E. (1964). Feynman integrals and the Schr´odinger equation, *J. Math. Phys.* **5**, p. 33.
[236] Nemytskii, V. and V., S. (1960). *Qualitative Theory of Ordinary Differential Equations* (Princeton University Press, Princeton).
[237] Newton, T. and Wigner, E. P. (1949). Localized states for elementary systems, *Rev. Mod. Phys.* **21**, p. 400.
[238] Nielsen, M. V. and Chuang, I. L. (2000). *Quantum computation and quantum information* (Cambridge University Press, Cambridge).
[239] Nolan, J. P. (2005). *Stable Distributions: Models for Heavy tailed Data* (Birkhauser, Boston, MA).
[240] Ord, G. N. (1992a). Classical analog of quantum phase, *Int. J. Theor. Physics* **31**, p. 1177.
[241] Ord, G. N. (1992b). A reformulation of the Feynman chessboard model, *J. Stat. Phys.* **66**, p. 647.
[242] Ord, G. N. and McKeon, D. G. C. (1993). On the Dirac equation in $3+1$ dimensions, *Annals of Physics* **222**, p. 244.
[243] Ott, E. (2002). *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge).
[244] Ozawa, M. (2003). Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement, *Phys. Rev. A* **67**, p. 042105.
[245] Ozawa, M. (2005). Universal uncertainty principle in the measurement operator formalism, *J. Opt. B* **7**, p. 672.
[246] Pal, N., Jin, C. and Lim, W. K. (2005). *Handbook of Exponential and Related Distributions for Engineers and Scientists* (Taylor & Francis Group, New York, USA).
[247] Palmer, R. G. (1982). Broken ergodicity, *Adv. Phys.* **31**, p. 669.
[248] Paluš, M., Komárek, V., Hrnčíř, Z. and Štěrbová, K. (2001). Synchronization as adjustment of information rates: detection from bivariate time series, *Phys. Rev. E* **63**, p. 046211.
[249] Pawula, R. F. (1967). Approximation of the linear Boltzmann equation by the Fokker–Planck equation, *Phys. Rev.* **162**, p. 186.
[250] Pearl, J. (2009). *Causality* (Cambridge University Press, Cambridge, UK).
[251] Peng, C.-K., Havlin, S., Buldyrev, S., Stanley, H. E. and Goldberger, A. (1994). Mosaic organization of dna nucleotides, *Phys. Rev. E* **49**, p. 1685.
[252] Peng, C.-K., Havlin, S., Stanley, H. E. and Goldberger, A. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, *Chaos* **5**, p. 82.
[253] Petersen, K. (1983). *Ergodic Theory* (Cambridge University Press, Cambridge).
[254] Petz, D. (1986). Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* **23**, p. 57.
[255] Pollicott, M. and Yuri, M. (1998). *Dynamical systems and ergodic theory* (Cambridge University Press, Cambridge).

[256] Polyakov, A. M. (1987). *Gauge Fields and Strings* (Harwood, New York).
[257] Ram, E. and Sason, I. (2016). On Rényi entropy power inequalities, *IEEE Trans. Inf. Theory* **62**, p. 6800.
[258] Ramond, P. (2001). *Field Theory: A Modern Primer, 2nd edition* (Westview Press, New York, USA).
[259] Rao, C. R. (1987). Differential metrics in probaility spaces, in *Differential Geometry in Statistical Inference* (IMS–Lecture Notes, Hayward, California), pp. 119–238.
[260] Reed, M. and Simon, B. (1975). *Methods of Modern Mathematical Physics. Vol. XI* (Academic Press, New York, USA).
[261] Reinolds, A. M. (2003). Superstatistical mechanics of tracer-particle motions in turbulence, *Phys. Rev. Lett.* **91**, p. 084503.
[262] Renner, R. (2005). Security of quantum key distribution, *PhD thesis.* **ETH Zurich**.
[263] Rényi, A. (1970). *Probability Theory* (North-Holland, Amsterdam).
[264] Rényi, A. (1976). *Selected Papers of Alfréd Rényi* (Akademia Kiado, Budapest).
[265] Resnic, S. I. (1971). Asymptotic location and recurrence properties of maxima sequence of random variables defined on a markov chain, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **18**, p. 197.
[266] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion* (Springer, Berlin).
[267] Riazanov, G. V. (1958). The Feynman path integral for Dirac equation, *Sov. Phys. JETP* **6**, p. 1107.
[268] Riesz, M. (1926). Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires, *Acta. Math.* **49**, p. 465.
[269] Risken, H. (1996). *The Fokker–Planck Equation Methods of Solution and Applications* (Springer, Stuttgart).
[270] Rivers, R. J. (1987). *Path integral methods in quantum field theory* (Cambridge Univ. Press, Cambridge, UK).
[271] Robertson, H. P. (1929). The uncertainty principle, *Phys. Rev.* **34**, p. 163.
[272] Rockafellar, R. T. (1972). *Convex Analysis* (Princeton University Press, Princeton).
[273] Rothstein, J. (1971). *Informational Generalization of Entropy In physics* (Cambridge University Press, London), ed. Bastin, T.
[274] Rovelli, C. and Speziale, S. (2003). Reconcile Planck-scale discreteness and the Lorentz–Fitzgerald contraction, *Phys. Rev. D* **67**, p. 064019.
[275] Samko, S. G., Kilbas, A. A. and Marichev, O. I. (1993). *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach, New York, USA).
[276] Satin, F. and Salasnich, L. (2002). Multiparameter generalization of nonextensive statistical mechanics, *Phys. Rev.* **E65**, p. 035106.
[277] Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions* (Cambridge Univ. Press, Cambridge, UK).
[278] Scardigli, F., Blasone, M., Luciano, G. and Casadio, R. (2018). Modified Unruh effect from generalized uncertainty principle, *Eur. Phys. J. C* **78**,

p. 728.
[279] Scardigli, F., Lambiase, G. and Vagenas, E. (2017). GUP parameter from quantum corrections to the Newtonian potential, *Phys. Lett. B* **767**, p. 242.
[280] Schreiber, T. (2000). Measuring information transfer, *Phys. Rev. Lett.* **85**, p. 461.
[281] Schrödinger, E. (1930). About Heisenberg uncertainty relation, *Sitzungsber. Preuss. Akad. Wiss.* **24**, p. 296.
[282] Schulman, L. S. (1981). *Techniques and Applications of Path Integration* (Wiley, New York).
[283] Schuster (1995). *Deterministic Chaos: An Introduction, 3rd ed.* (Wiley, New York).
[284] Schweber, S. (1986). Feynman and the visualization of space-time processes, *Rev. Mod. Phys.* **58**, p. 449.
[285] Shannon, C. E. (1948a). A mathematical theory of communication I, *Bell Syst. Tech. J.* **27**, p. 379.
[286] Shannon, C. E. (1948b). A mathematical theory of communication II, *Bell Syst. Tech. J.* **27**, p. 623.
[287] Shannon, C. E. (1993). *Collected Papers* (IEEE Press, New York), eds. Sloane, N. J. A. and Wyner, A. D.
[288] Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication* (University of Illinois Press, New York).
[289] Sharma, B. D. and Mittal, D. P. (1975). New non-additive measures of entropy for discrete probability distributions, *J. Math. Sci.* **10**, p. 28.
[290] Shore, J. E. and Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Transactions on Information Theory* **26**, p. 26.
[291] Sinai, Y. G. E. (1997). *Dynamical Systems II. Ergodic Theory with Applications to Dynamical Systems and Statistical Mechanics* (Springer-Verlag, Berlin).
[292] Smolin, L. (2008). Could deformed special relativity naturally arise from the semiclassical limit of quantum gravity? *arXiv: 0808.3765 [hep-th]* .
[293] Smolin, L. (2010). On limitations of the extent of inertial frames in noncommutative relativistic spacetimes, *arXiv: 1007.0718 [gr-qc]* .
[294] Snyder, H. S. (1947). Quantized space-time, *Phys. Rev.* **71**, p. 68.
[295] Stanley, H. E. (1971). *Introduction to phase transitions and critical phenomena* (Clarendon Press, Qxford).
[296] Stanley, H. E. and Meakin, P. (1988). Multifractal phenomena in physics and chemistry, *Nature* **335**, p. 405.
[297] Stroock, D. W. and Varadhan, S. R. S. (1973). Limit theorems for random walks on lie groups, *Sankhya Ser. A* **35**, p. 277.
[298] Stueckelberg, E. C. G. (1941). The stable wave packet and uncertainty, *Helvetica Physica Acta* **14**, p. 588.
[299] Stueckelberg, E. C. G. (1942). The mechanics of point particles in the theory of relativity and the quantum theory, *Helvetica Physica Acta* **15**, p. 23.
[300] Swanson, M. S. (1993). Phase-space anomalies and canonical transforma-

tions, *Phys. Rev.* **A47**, p. 2431.
[301] Synge, J. L. (1971). Geometrical approach to the Heisenberg uncertainty relation and its generalization, *Proc. R. Society London* **A325**, p. 151.
[302] Szilárd, L. (1929). On the reduction of entropy in a thermodynamic system by the intervention of intelligent beings, *Zeit.Phys.* **53**, p. 840.
[303] Tagliani, A. (1998). Inverse two-sided Laplace transform for probability density functions, *J. Comp. Appl. Math.* **90**, p. 157.
[304] Talker, P. and Weber, R. (2000). Power spectrum and detrended fluctuation analysis: application to daily temperatures, *Phys. Rev. E.* **62**, p. 150.
[305] Theiler, J. (1990). Estimating fractal dimension, *J. Opt. Soc. Amer.* **A 7**, p. 1055.
[306] Thorin, O. (1939). An extension of the convexity theorem of M. Riesz. *Kungl. Fysiogr. Sällsk. Lund Förh.* **8**, p. 14.
[307] Thurner, S., Hanel, R. and Klimek, P. (2018). *Introduction to the Theory of Complex Systems* (Oxford University Press, Oxford, UK).
[308] Toda, M., Kubo, R. and Saito, N. (1992). *Statistical Physics I: Equilibrium Statistical Mechanics* (Springer-Verlag, London, UK).
[309] Touchette, H. (2002). When is a quantity additive, and when is it extensive? *Physica A* **305**, p. 84.
[310] Touchette, H. and Beck, C. (2005). Asymptotics of superstatistics, *Phys. Rev.* **E71**, p. 016131.
[311] Tsallis, C. (1988). Possible generalization of Boltzmann–Gibbs statistics, *J. Stat. Phys.* **52**, p. 479.
[312] Tsallis, C. (1999). Nonextensive statistics: Theoretical, experimental and computational evidences and connections, *Braz. J. Phys.* **29**, p. 1.
[313] Tsallis, C. (2002). Nonextensive statistical mechanics: a brief review of its present status, *Annals Braz. Ac. Sci.* **74**, p. 393.
[314] Tsallis, C. (2009). *Introduction to Nonextensive Statistical Mechanics; Approaching a Complex World* (Springer, New York).
[315] Tsallis, C. (2012). *http://tsallis.cat.cbpf.br/biblio.htm* .
[316] Tsallis, C., Anteneodo, C., Borland, L. and Osorio, R. (2003). Nonextensive statistical mechanics and economics, *Physica A* **89**, p. 89.
[317] Tsallis, C., Anteneodo, C. and Querios, S. M. D. (2005). Power-law distributions in economics: A nonextensive statistical approach, *Proc. SPIE 5848* **89**, p. 151.
[318] Tseng, C.-Y. and Caticha, A. (2002). Yet another resolution of the gibbs paradox: an information theory approach, *AIP Conference Proceedings* **617**, p. 331.
[319] Umarov, S., Tsallis, C., Gell-Man, M. and Steinberg, S. (2010). Generalization of symmetric α-stable lévy distributions for $q > 1$, *J. Math. Phys.* **51**, p. 033502.
[320] Umarov, S., Tsallis, C. and Steinberg, S. (2008). On a q-central limit theorem consistent with nonextensive statistical mechanics, *Milan J. Math.* **76**, p. 307.
[321] Vedral, V. (2001). The role of relative entropy in quantum information theory, *Rev. Mod. Phys.* **74**, p. 197.

[322] Vedral, V. and Plenio, M. B. (1998). Entanglement measures and purification procedures, *Phys. Rev. A* **57**, p. 1619.
[323] Vedral, V., Plenio, M. B., Rippin, M. A. and Knight, P. L. (1997). Quantifying entanglement, *Phys. Rev. Lett.* **78**, p. 2275.
[324] Vidal, G. and R., T. (1999). Robustness of entanglement, *Phys. Rev. A* **59**, p. 141.
[325] Vilenkin, A. and Ford, L. H. (1982). Gravitational effects upon cosmological phase transitions, *Phys. Rev. D* **26**, p. 1231.
[326] Voit, J. (2003). *The Statistical Mechanics of Financial Markets* (Springer, Amsterdam, Netherlands).
[327] Wallace, D. L. (1958). Asymptotic approximations to distributions, *Ann. Math. Stat.* **29**, p. 635.
[328] Wei, L. (2019). On the exact variance of tsallis entanglement entropy in a random pure state, *Entropy* **21**, p. 539.
[329] Weiss, U. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (John Wiley and Sons, Inc., New York).
[330] Wiener, N. (1923). Differential-space, *J. Math. & Phys.* **2**, p. 132.
[331] Wiener, N. (1956). The theory of prediction, in *Modern Mathematics for Engineers* (McGraw-Hill, New York, USA), p. 23.
[332] Wigner, E. P. and Yanase, M. M. (1963). Information contents of distributions, *Proc. N. A. S.* **49**, p. 910.
[333] Wilk, G. and Wlodarzyk, Z. (2000). Interpretation of the nonextensivity parameter q in some applications of Tsallis statistics and Lévy distributions, *Phys. Rev. Lett.* **84**, p. 2770.
[334] Witten, T. A. and Sander, L. M. (1981). Diffusion-limited aggregation, a kinetic critical phenomenon, *Phys. Rev. Lett.* **47**, p. 1400.
[335] Zinn-Justin, J. (2007). *Phase transitions and renormalization group* (Oxford University Press, Oxford, UK.).
[336] Zolotarev, V. M. (1957). Mellin–Stieltjes transforms in probability theory, *Teor. Veroyatnost. i Primenen.* **2**, p. 444.

Index