On Non-Extensive Entropies: With Applications in Stochastic Dynamics and Information Theory

Jesús Fuentes

Department of Physics

University of Guanajuato
División de Ciencias e Ingenierías, Campus León

A thesis submitted for the degree of

Doctor in Physics

August 2021

1. Reviewer: Dr. Nana Cabo Bizet

2. Reviewer: Dr. Alberto Diez Tejedor

3. Reviewer: Dr. Hugo García-Compeán

4. Reviewer: Dr. José Luis López Picón

5. Reviewer: Dr. Octavio Obregón

6. Reviewer: Dr. Carlos Wiechers

Day of the defence:

Signature from head of PhD committee

Abstract

Among the generalised measures of entropy, there is a special class of measures whose functional dependence dismisses all free parameters, but instead relies exclusively on probability. From this class, we will pay attention to full-stable measures of entropy having a well defined thermodynamic limit, provided these attributes are necessary for physical observables to be recovered from entropy. To our knowledge, there are only two generalised entropies fulfilling these requirements. Then we investigate their basic mathematical aspects as well as their impact on physics, information and computer sciences. We will prove formally such entropies converge asymptotically to the Boltzmann-Gibbs measure, whereas they induce a generalised classification of entropies. We study the consequences these entropies convey in diffusion and transport phenomena, which leads us to derive master equations out of equilibrium. Interestingly, our master equations adopt a similar structure to some chemotaxis-aggregation models studied in biology. Further, given that entropy is at the interface between statistical mechanics and information theory, we propose a non-extensive information theory, where data compression and channel capacities are improved, in relation to Shannon's formulation, in a scenario of high probabilities. Finally, we bring this non-extensive information theory in its algorithmic counterpart to obtain generalisations to Kolmogorov's statistical complexity.

To my father

Acknowledgements

I want to especially thank Prof. Octavio Obregón for leading me wisely through my doctoral studies. I treasure the pragmatism of his teachings, the encourage he pushed on me to imagine rather than calculate, and the freedom he gave me to develop my own ideas.

I much appreciate the discussions and feedback from Dr. José Luis López and Dr. Nana Cabo, who provided input into some of the findings presented in this dissertation. They taught me a lot.

The observations, advice and screenings of Dr. Alberto Diez, Dr. Carlos Wiechers and Dr. Hugo García-Compeán are highly valued. Thank you for your help, despite the workload on your calendars.

My gratitude is expressed to Norma López for her kind assistance. I put her patience into the test.

I am grateful to the University of Guanajuato and its warmth community.

I thank CONACyT for the funding that made it possible for me not to compromise my performance during the doctoral programme.

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

1

Introduction

In the 1860s, the German physicist Rudolf Clausius coined the term *entropy* to denote the amount of heat loss upon to performing work[20]. He claimed when a physical system passes from state A to B, the heat absorbed $\mathrm{d}Q$ from its surroundings induces a change of entropy ΔS such that

$$\Delta S = \int_A^B \frac{\mathrm{d}Q}{T} \,,$$

where T is the absolute temperature of the surroundings of the system at the moment of the transition. Insofar as the process is irreversible, entropy will always increase, while remaining constant within each reversible process.

Years later, Boltzmann developed a fascinating idea about entropy[6]. Taking a statistical view, he claimed the entropy of an ideal gas is proportional to the logarithm of the number of microstates that the gas might occupy. His ideas, together with those of Gibbs and Maxwell, gave rise to the statistical formulation of thermodynamics, where the basic problem is to obtain the distribution of energy—and other observables—over a defined number of identical systems.

In quantum mechanics, the extension of classical Boltzmann-Gibbs entropy was given by von Neumann[79]. In this context, entropy estimates the deviation of the system from a pure state. So while the system will have a vanishing entropy for a pure state, entropy will increase as the quantum interference is erased, since a quantum of transitional energy is required from state A to state B.

A crucial development came during the 1940s, when Claude Shannon[66] formulated a mathematical theory to quantify and compress information—which is still essential

to our daily digital communication processes. He discovered the minimum compression ratio a data communication would achieve without losing information is estimated by the entropy itself. Not merely has this been fundamental in electrical engineering and computer sciences, but moreover, since Shannon's entropy and that of Boltzmann-Gibbs own the same mathematical structure, there exists a natural connection between information theory and statistical mechanics.

Today, the notion of entropy plays a central role in maximising those probability distributions that describe the microscopic behaviour of generic statistical systems. However, while standard distributions successfully represent the thermodynamics of weakly interactive systems involving a large number of microstates, this is not the case for complex systems, i.e. systems having a high degree of interaction between components. In these cases it is necessary to introduce *generalised entropies* [1,43,52,62,68,72].

To take into account the thermodynamics of complex systems outside equilibrium, various measures of entropy depending on probability and free parameters, have been suggested directly or indirectly. For example, it was well shown in Ref. [24] that auto-similar structures derived from scaling properties in Yang-Mills theories behave as fractals, from which a non-extensive statistical pattern is obtained leading to non-extensive entropies that convey complexity.

In recent years, by following the superstatistics formalism [4], it was proposed in Ref. [56] a pair of non-extensive entropies, S_+ and S_-, which depend only on the probability and having the particularity of looking like Boltzmann-Gibbs entropy in the first approximate order. While these generalised statistics resemble systems slightly outside the equilibrium, there is a temporal local equilibrium in each of the cells that compose the system. In fact, this type of entropies results in long-term stationary states taking into account the spatiotemporal fluctuations of intensive quantities, such as the inverse temperature. Physically, this delivers the response of a system to attractive or repulsive effective interactions [8,58].

In this work we focus on the analysis of the non-extensive entropies S_+ and S_-, we shall look at their mathematical attributes and their physical consequences in connection with information theory.

Along our exposition, all the systems we consider are modestly outside the thermodynamic equilibrium, but having long-term stationary states so that the amplitudes of the fluctuations will remain small even when they are subject to small variations

in the initial conditions. This stable behaviour also grants universality to the scheme, that is, the corrections made to the standard Boltzmann factor must be almost the same for all superstatistics if a well defined thermodynamic limit is demanded. Those corrections to the Boltzmann-Gibbs entropy that S_+ and S_- convey, are monotonically subdominant in both cases. We have to remark this is an interesting attribute since such tendency indicates S_+, S_- and S_{BG} will coincide asymptotically, thus retrieving the standard theory when S_+ or S_- encompass a large number of accessible states, whereas for a contracted number of them, the modest differences between them will lead us to discuss worth applications.

Our work consists of four independent chapters, which are the result of our own research in the area of non-extensive entropies, plus a chapter devoted to the conclusions and perspectives. We summarise our exposition's structure as follows.

Chapter 2 We discuss the mathematical attributes of S_+ and S_-. We show these entropies are asymptotically equivalent to the Boltzmann-Gibbs entropy and give a formal proof of their stability. We also discuss how the corresponding entropic forms of S_+ and S_- bring a generalised family of entropies, forthwith passing to a hierarchy of measures of entropy classified according to their scaling exponents. Our results were published in Ref.[9].

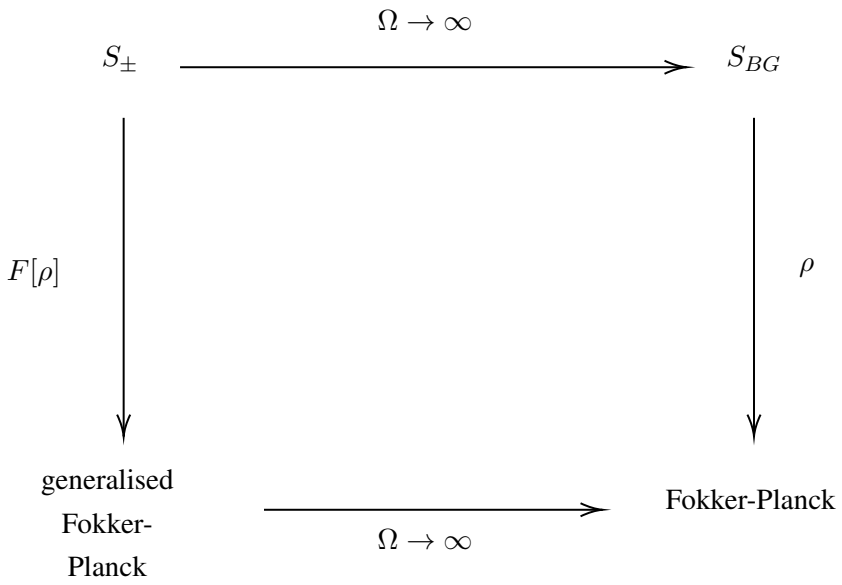
Chapter 3 As part of the framework elaborated in Chapter 2, we implement an entropy method to derive generalised Fokker-Planck equations. The motivations are simple: If classical Boltzmann-Gibbs entropy leads to the mean-field Fokker-Planck equation, which equations would emerge from generalised measures of entropy?

We shall limit our discussion to the generalised Fokker-Planck equations obtained from entropies S_+ and S_-. Both equations will inherit from these entropies the asymptotic behaviour, hence, in a scenario where the number of available states is much greater than the particles involved, our equations' description will coincide with that given by the original master equations, see the diagram below.

Some numerical experiments are studied, showing the anomalous diffusion phenomenon carried by our models and the transient behaviour observed in electron transport. Furthermore, it is worth to mention that the kind of Fokker-Planck

equations derived here emerge, as well, in several stochastic processes studied in biological sciences, e.g. we compare our equations with a chemotaxis-aggregation model.

The results reported in this chapter, were published in Ref.[31].



Chapter 4 In this chapter, we examine the links between statistical mechanics and information theory, of which entropy is a common node. We show that entropies S_+ and S_- (which will be referred as H^+ and H^- in the field of information) improve certain information processes, such as data compression and channel capacities, in relation with Shannon's entropy.

We are also developing a variational method for finding optimal codeword lengths, which allow the formulation of generalised noiseless coding theorems related to generalised measures of entropy. The outcomes referred to in this chapter were reported in Ref.[29].

Chapter 5 In another application of statistical mechanics to information theory, we move the formulation of superstatistics to the domain of complexity and algorithmic theory of information. This brings us to propose possible generalisations for the complexity of Kolmogorov and to foresee an infinite number of measures of complexity. Still, not all complexity measures are likely to be computationally affordable, due to stability reasons. So our focus is on the complexity measures

associated with H^+ and H^-. As a result of our research, findings in this area were reported in Ref.[30].

Chapter 6 Finally, we conclude our work with a general and forward-looking discussion in the field of non-extensive entropies.

2

Measures of Entropy

Entropy is a node where various edges converge. It is a fundamental quantity in classical thermodynamics, the interface between the microscopic and the macroscopic world in statistical physics and the key concept in information theory. In modern mathematics, this notion is widely utilised from topological geometry to combinatorics. Although, in the same way, it is frequently encountered in computer sciences, chemistry, biology, economy and sociology. The use of entropy is so broadly diverse that trying to state a universal definition would not only be unfortunate but surely incorrect.

In this chapter, we focus on the statistical treatment of entropy, whose application to physics is direct and is equivalent to that of information theory—that we shall survey in the second part of this work. In what follows, we introduce generalised entropies. However, we emphasise those functional ones that depend solely on the probability distribution. Further, the functionals of our interest are also physically congruous with the standard formulation of thermostatistics.

2.1 Generalised Entropies

In the Boltzmann formulation of thermostatistics entropy S is a quantity both additive and extensive. By additive we mean, if a thermodynamic system with Ω states is divided into N subsystems, each one associated with a number $\Omega_1, \Omega_2, \ldots, \Omega_N$ of states, then the total entropy of the system satisfies the functional equation $S(\Omega_1\Omega_2\cdots\Omega_N) = S(\Omega_1) + S(\Omega_2) + \cdots + S(\Omega_N)$. Conversely, if Ω is composed as $\Omega_{1+2+\ldots+N}$ such that $S(\Omega_{1+2+\ldots+N}) = S(\Omega_1) + S(\Omega_2) + \cdots + S(\Omega_N)$, we say that the entropy is extensive,

inasmuch as the N subsystems do not interact with each other. Clearly, additivity and extensiveness coincide whenever $\Omega_{1+2+\cdots+N} = \Omega_1\Omega_2\cdots\Omega_N$. For these reasons, in standard thermodynamics we say the value of S depends on the amount of material.

There are, however, circumstances in which the former does not hold. As proof, the condition $S(\Omega_{1+2+\cdots+N}) = S(\Omega_1) + S(\Omega_2) + \cdots + S(\Omega_N)$ is immediately broken when the system possesses inner correlations, thus implying the individual number of states overlaps at some point. Although even in the absence of interactions the entropy production can obey a non-extensive rule. That is the case of charged fields at fixed temperature in a free bosonic quantum field theory, which endow the geometric entropy [10,41,42] with nonextensive terms [32].

In general, therefore, there is a need to appraise entropies beyond the standard definition due to Boltzmann and Gibbs,

$$S_{BG} = -k_B \sum_j^{\Omega} p_j \ln p_j, \quad 0 \leq p_j \leq 1 \,, \tag{2.1}$$

k_B is a universal constant, that from now on we equal to unit, and $P = \{p_1, p_2, \ldots, p_\Omega\}$ is a probability distribution associated with a thermodynamic scenario. S_{BG} is interpreted as a logarithmic measure that estimates the number of microstates that can be occupied according to a distribution P. Yet this estimation will vary depending on the functional measure that characterises the entropy, as seen from the following definition.

Definition 1. *A generalised entropy defines as*

$$S = \sum_j^{\Omega} g(p_j) \,, \tag{2.2}$$

where g is a monotonic increasing function known as entropic form.

Remark 1. *The form (2.2) is a special case of $S = \phi\left(\sum_j^{\Omega} g(p_j)\right)$, where ϕ is a Nagumo-Kolmogorov function. We also will survey entropies of this type, nonetheless, for the sake of simplicity we shall consider (2.2) as the prototype definition of entropy.*

Remark 2. *If $g(x) = -x\ln x$, then the entropy (2.1) is automatically resembled. Furthermore, a particular attribute of S_{BG} is that it completely satisfies the following definition.*

Definition 2. *The Shannon-Khinchin (SK) axioms for entropy S read [44]:*

SK1 *$S : P \to \mathbb{R}^+$ is continuous in P.*

SK2 *S takes its largest value for a uniform distribution, $p_j = 1/\Omega$ for all j.*

SK3 *S is expandable, $S(p_1, p_2, \ldots, p_{\Omega-1}, 0) = S(p_1, p_2, \ldots, p_{\Omega-1})$.*

SK4 *S is additive, $S(PQ) = S(P) + S(Q)$, where $P = \{p_1, p_2, \ldots, p_\Omega\}$ and $Q = \{q_1, q_2, \ldots, q_{\Omega'}\}$.*

While some of these axioms may be violated by generalised entropies, we shall limit our interest to only those entropies that violate SK4 while fulfilling SK1-SK3 thoroughly. As we will see, by looking at non-extensive entropies with these requirements, it is possible to construct equivalence classes that ultimately allow stratifying entropies into hierarchies.

Our whole discussion centres in the non-extensive entropies

$$S_+ = \sum_j^\Omega \left(1 - p_j^{p_j}\right), \quad \text{and} \quad S_- = \sum_j^\Omega \left(p_j^{-p_j} - 1\right) , \tag{2.3}$$

which are of the form (2.2) and fulfil SK1-SK3 for any distribution P, whereas SK4 is partially violated. For the sake of illustrating this latter fact, it is convenient to expand the two entropies in powers of $p_j \ln p_j$ as:

$$S_+ = -\sum_j^\Omega \left(p_j \ln p_j + \frac{1}{2!}(p_j \ln p_j)^2 + \frac{1}{3!}(p_j \ln p_j)^3 + \cdots\right) ,$$

and

$$S_- = -\sum_j^\Omega \left(p_j \ln p_j - \frac{1}{2!}(p_j \ln p_j)^2 + \frac{1}{3!}(p_j \ln p_j)^3 - \cdots\right) ,$$

in both cases, the first term is the Boltzmann-Gibbs entropy, which leads the series. The higher-order terms become negligible if P is very low, i.e. if the number of available microstates Ω is much greater than the number of particles. In thermodynamics, this scenario reproduces at high temperatures. Therefore, in the thermodynamic limit, entropies (2.3) resemble (2.1) and SK4 is satisfied as well. In other words, entropies S_+, S_- and S_{BG} belong to the same equivalence class.

Furthermore, focusing on a microcanonical configuration (E, V, N) for which the functionals S_+ and S_- attain their maximum at $p_j = 1/\Omega$ for every j, we expressly have:

$$S_\pm = \pm\Omega \mp \Omega^{1\mp 1/\Omega} ,$$

expanding in series, we get

$$S_{\pm} = \sum_{j=1}^{\infty} (\mp 1)^{j+1} \frac{\ln^j \Omega}{j! \Omega^{j-1}} \ .$$

As we have previously mentioned, the contribution of the higher-order terms becomes negligible as Ω grows since $\ln^j \Omega < \Omega^{j-1}$, for $j > 1$. Hence the quotient $\ln^j \Omega / \Omega^{j-1} \sim 0$ for $\Omega \gg 1$. Actually, since $S_{BG} = \ln \Omega$ as for an equipartition configuration, we get

$$S_{\pm} = \sum_{n=1}^{\infty} (\mp 1)^{n+1} \frac{S_{BG}^n}{n! \exp[(n-1) S_{BG}]} \ ,$$

which is graphically presented in Figure 2.1. Note the subtle differences conveyed by S_+ and S_- with respecto to S_{BG} when the number of available states Ω is approximately less than 400 ($\ln 400 \approx 6$).

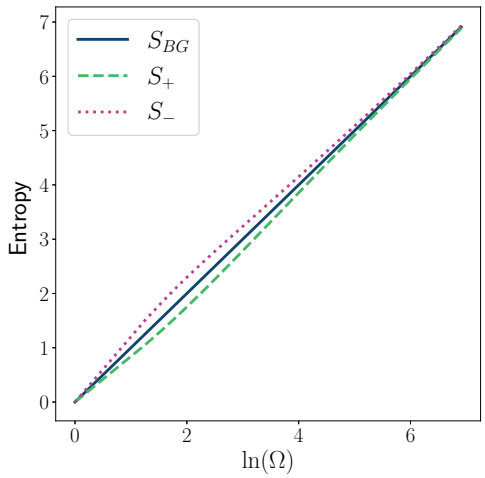


Figure 2.1: Thermodynamic compatibility of S_+ and S_- - Behaviour of S_+, S_- and S_{BG} in terms of a uniform distribution $p_j = 1/\Omega$, for all j. The three entropies coincide as the number of available microstates grows.

The functionals (2.3) were originally proposed by Obregón[56] within the superstatistics framework of effective Boltzmann factors[4]:

$$B(E) = \int_0^{\infty} \mathrm{d}\beta \, f(\beta) e^{-\beta E}, \quad \beta \in \mathbb{R} \ , \tag{2.4}$$

where f is an arbitrary distribution, E is the internal energy and β is a Lagrange multiplier. If the integral is computable, it is easy to find the generalised entropy with the aid of formula[76]

$$g(x) = \int_0^x \mathrm{d}y \, \frac{\alpha + E}{1 - E/E^*} \, , \tag{2.5}$$

the energy E, with minimum value E^*, is obtained by inverting $B(E)$ and α is a constant for determining from the condition $g(1) = 0$.

As for entropies S_+ and S_-, the corresponding distributions that feed the integral in (2.4) are Gamma-like distributions characterising a global fluctuation of temperatures with initial datum β_0 and a parameter p_j, eventually identified with the probability, namely

$$f^{\pm}_{p_j}(\beta) = \frac{1}{\beta_0 p_j \Gamma\left(\frac{1}{p_j}\right)} \left(\frac{\beta}{\beta_0}\frac{1}{p_j}\right)^{\frac{\pm 1 - p_j}{p_j}} \exp\left(-\frac{\beta}{\beta_0 p_j}\right) \, ,$$

which produce the pair of effective Boltzmann factors

$$B^{\pm}_{p_j}(E) = (1 \pm p_j \beta_0 E)^{\mp \frac{1}{p_j}} \, . \tag{2.6}$$

These generalised Boltzmann factors relate to systems whose cells are in local equilibrium—β_0 remains approximately fixed in each cell—while the system slightly deviates from global equilibrium. For instance, consider the expansions

$$B^{+}_{p_j}(E) = e^{-\beta_0 E}\left(1 + \frac{1}{2}(\beta_0 E)^2 p_j + \frac{1}{24}(\beta_0 E)^3(3\beta_0 E - 8)p_j^2 + \cdots\right) \, , \tag{2.7}$$

and

$$B^{-}_{p_j}(E) = e^{-\beta_0 E}\left(1 - \frac{1}{2}(\beta_0 E)^2 p_j + \frac{1}{24}(\beta_0 E)^3(3\beta_0 E - 8)p_j^2 - \cdots\right) \, , \tag{2.8}$$

observe that as the probability approaches zero the contribution of the higher-order terms tends to vanish. A similar behaviour occurs for high energies, where the number of accesible states is much greater than the number of particles, see Figure 2.2.

In the next chapter, we are to survey some of the physical consequences conveyed by the entropies (2.3) in stochastic dynamical systems. Yet, in recent studies, it has been discussed their relevance in configurations with few available microstates[57]. E.g. a study in molecular dynamics[33] shows that the entropies S_+ and S_- describe the presence of effective forces that would be negligible in a large-scale layout. The implications in Bose-Einstein condensation have been discussed with detail in Ref.[58]. And

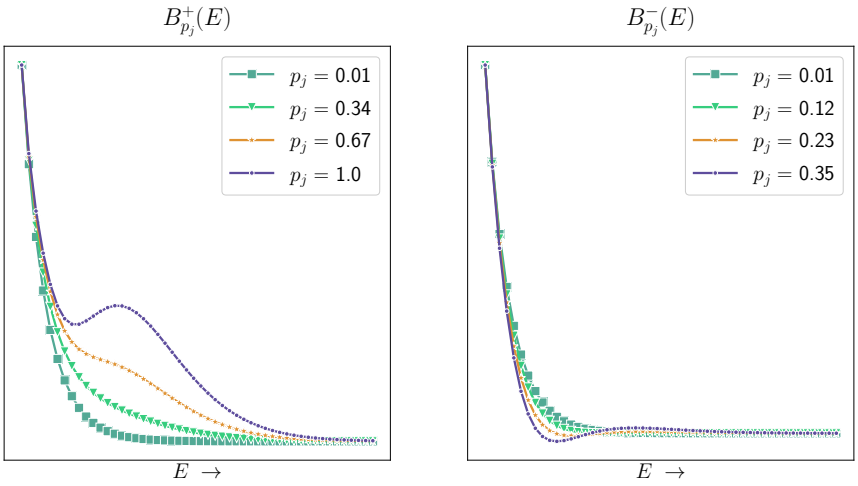


Figure 2.2: Effective Boltzmann factors - The figure shows the generalised factors in (2.7) and (2.8) for fixed β_0 and varying p_j and E. As the probability decreases the standard Boltzmann factor is resembled.

the quantum counterparts of entropies (2.3) have been obtained in Ref.[8] by generalising the replica trick—a common technique to compute the partition function in spin glasses.

2.2 Boltzmann Equivalence Class

As we have suggested, there is an asymptotic equivalence shared by entropies S_+, S_- and S_{BG}, which is evident from the series expansions presented in the previous section. Nonetheless, an asymptotic analysis conduces to a more formal approach and indeed allows classifying measures of entropy in terms of scale exponents. Such a formalism was proposed by Hanel & Thurner in Ref.[34]. In what follows, we shall discuss the key ideas behind their analysis to show that S_+, S_- and S_{BG} belong to the same asymptotic, equivalence class.

Definition 3. *Any generalised entropy satisfying the axioms SK1-SK3 is thoroughly characterised by a pair of scale exponents (c, d) pertaining, respectively, to the asymptotic laws*

$$\lim_{\Omega\to\infty} \lambda \frac{g\left(\frac{1}{\lambda\Omega}\right)}{g\left(\frac{1}{\Omega}\right)} = \lambda^{1-c}, \quad 0 < c \leq 1, \quad \lambda \in \mathbb{R}\,, \tag{2.9}$$

and

$$\lim_{x\to 0}\frac{g\left(x^{1+a}\right)}{x^{ac}g(x)}=(1+a)^d,\quad a\in\mathbb{R}\,, \tag{2.10}$$

the constants a,λ *are arbitrary.*

It is easy to verify that entropies S_+, S_- and S_{BG} are characterised by the same pair of scale exponents $(c,d)=(1,1)$. We start with the case of S_{BG} and the law (2.9),

$$\lim_{\Omega\to\infty}\frac{\frac{1}{\Omega}\ln\frac{1}{\lambda\Omega}}{\frac{1}{\Omega}\ln\frac{1}{\Omega}}=\lim_{\Omega\to\infty}\frac{\ln\frac{1}{\lambda\Omega}}{\ln\frac{1}{\Omega}}=1=\lambda^{1-c}\,,$$

which means $c=1$. The same for Eq. (2.10),

$$\lim_{x\to 0}\frac{x^{1+a}\ln x^{1+a}}{x^{1+a}\ln x}=1+a=(1+a)^d\,,$$

thus concluding S_{BG} belongs to the equivalence class $(c,d)=(1,1)$. As for S_+ and Eq. (2.9), this yields

$$\lim_{\Omega\to\infty}\lambda\frac{1-\left(\frac{1}{\lambda\Omega}\right)^{\frac{1}{\lambda\Omega}}}{1-\left(\frac{1}{\Omega}\right)^{\frac{1}{\Omega}}}=1=\lambda^{1-c}\,,$$

therefore $c=1$. Now, applying Eq. (2.10)

$$\lim_{x\to 0}\frac{1-\left(x^{1+a}\right)^{x^{1+a}}}{x^{ac}(1-x^x)}=\lim_{x\to 0}\frac{1-\left(x^{1+a}\right)^{x^{1+a}}}{x^{a}(1-x^x)}=1+a=(1+a)^d\,,$$

hence $d=1$. We proceed analogously in the case of S_- to obtain the same exponents $(c,d)=(1,1)$. In addition, S_+, S_- and S_{BG} share another interesting attribute, they accept the same continuum representation, as it is discussed in Appendix A.

Some comments are in order. Since $c=1$, it is necessary $d\geq 0$ to have SK2 fulfilled, which is satisfied inasmuch as g is a convex function. In the case of entropy S_{BG} this is a well known fact, but we shall prove that this condition is also satisfied by S_+ and S_-.

The convexity of S_+ and S_- is easily demonstrated as long as the condition

$$tS(p)+(1-t)S(p')\geq S(tp+(1-t)p'),\quad 0\leq t\leq 1\,,$$

is complied.

Beginning with S_+, we get

$$(1-t)\sum_j^{\Omega}p_j^{\prime p_j'}-t\sum_j^{\Omega}p_j^{p_j}\geq\delta\,,$$

such that

$$\delta \geq -\sum_{j}^{\Omega} \left[p'_j + t(p_j - p'_j)\right]^{p'_j + t(p_j - p'_j)} \ .$$

Given that inequalities must hold for every p_j and p'_j, the following is true:

$$(1-t)p_j^{\prime p'_j} > tp_j^{p_j} \quad \text{and} \quad \delta \leq 0 \ ,$$

therefore we have

$$\delta \leq (1-t)p_j^{\prime p'_j} - tp_j^{p_j} \leq 0 \ ,$$

and S_+ is convex. Ditto for S_-, we get

$$tp_j^{-p_j} + (1-t)p_j^{\prime -p'_j} \geq \left[t'p_j + (1-t')p'_j\right]^{-t'p_j - (1-t')p'_j} \ .$$

The right-hand member attains zero at $t' = p'_j/(p'_j - p_j)$ whereas the left-hand member at $t = -p_j^{p_j}/(p_j^{\prime p'_j} - p_j^{p_j})$, that is $t' > t$, hence S_- is convex. That concludes our proof.

2.3 An Extended Classification for Entropies

Despite S_+ and S_- are generalisations of S_{BG}, the equivalence class $(c, d) = (1, 1)$ reveals the three entropies relate to the same asymptotic behaviour. What we are to discuss below is a generalisation of the formulation introduced by Hanel & Thurner in Ref.[34]. While the authors' classification for entropy depends upon the Boltzmann generator $\ln x$, we consider the corresponding generators of S_+ and S_- instead. Accordingly, we shall introduce the functional generators:

$$\gamma_+(x) = 1 - x^x, \quad \text{and} \quad \gamma_-(x) = x^{-x} - 1 \ . \tag{2.11}$$

Definition 4. *For any pair of scaling exponents (c, d), there exists a factor*

$$\sigma = \frac{1}{1 + (d-1)c} \ ,$$

and a characteristic function $\phi(x)$, such that the functional generators $\gamma_\pm(x)$ in (2.11) define the set of universal entropic forms

$$g_\pm(x; c, d) = \sigma \left(\phi(x) - \Gamma[d+1, \alpha_\pm(x; c)]e^r\right) \ , \tag{2.12}$$

where

$$\alpha_\pm(x; c) = 1 - cW(-\gamma_\pm(x)), \quad r > 0 \ ,$$

$\Gamma[\cdot,\cdot]$ *is the incomplete gamma function and* $W(x)$ *is the product logarithm (Lambert function).*

The function $\phi(x)$ in Def. 4 is determined by each entropic form $g_{\pm}$ and the corresponding exponents (c,d). In the case of entropy S_+, we have $(c,d)=(1,1)$ and $\sigma=1$, which conduces to

$$\alpha_+ = 1 - W(x^x - 1) \ ,$$

having fixed $r=1$. In turn, the corresponding characteristic function is

$$\phi(x) = 2\left(1 - x^x + \exp\left[W\left(x^x - 1\right)\right]\right) \ ,$$

finally, from (2.12), one gets

$$g_+(x,1,1) = 1 - x^x \ ,$$

and S_+ is directly recovered according to (2.2).

An analogous procedure takes place for S_-, whose characteristic function reads

$$\phi_-(x) = 2\left(x^{-x} - 1 + \exp\left[W\left(1 - x^{-x}\right)\right]\right) \ ,$$

once again, after simplifying the expression in (2.12), one obtains

$$g_-(x,1,1) = x^{-x} - 1 \ ,$$

and the entropy S_- is recovered.

Remark 3. *The formulation in Ref.*[34] *is recovered if the generators* (2.11) *are at first-order truncated. I.e.* $\gamma_{\pm}(x) = -x\ln x + O(-\frac{1}{n!}(x\ln x)^n)$*, which means*

$$W(-\gamma_{\pm}(x)) \to \ln x \ ,$$

leading to $g_{\pm}(\cdot) \to g(\cdot)$*, where*

$$g(x;c,d) = \sigma\left[\phi(x) - \Gamma[d+1, a(x,c)]e^r\right] \ , \tag{2.13}$$

and $a(x;c) = 1 - c\ln x$*,* $r > 0$*.*

Remark 4. *The entropic form* (2.13) *characterised by the equivalence class* $(c,d) = (1,1)$ *yields* $g(x;1,1) = -x\ln x$*, restoring the Boltzmann-Gibbs entropy*

$$-\sum_j^{\Omega} p_j \ln p_j \ .$$

	Entropy	c	d
p-entropy $-$	$S_-(P) = \sum_j^\Omega \left(p_j^{-p_j} - 1\right)$	1	1
p-entropy $+$	$S_+(P) = \sum_j^\Omega \left(1 - p_j^{p_j}\right)$	1	1
W-exponential	$S_e(P) = \frac{1-\sum_j^\Omega \exp[rW(p_j^{p_j}-1)]}{r-1}$	$0 < r < 1$	0
Boltzmann	$S_B(P) = -\sum_j^\Omega p_j \ln p_j$	1	1
Tsallis	$S_q(P) = \frac{1-\sum_j^\Omega p_j^q}{q-1}$	$0 < q < 1$	0
Kaniadakis	$S_\kappa(P) = -\sum_j^\Omega p_j \frac{p_j^\kappa - p_j^{-\kappa}}{2\kappa}$	$0 \leq 1-\kappa < 1$	0

Table 2.1: Generalised entropies of the form (2.2). These entropies are particular cases of (2.12) as for Def. 4. Note that $S_\pm(P)$ and $S_B(P)$ are asymptotically equivalent, they belong to the equivalence class $(c, d) = (1, 1)$.

We are now in a position to classify various entropies, based on either the Boltzmann generator $\ln x$ or the generators (2.11) as well as the equivalence class defined by (c, d). Some entropies of regular interest are shown in Table 2.1.

Worth mentioning the W-exponential entropy—an outcome proper for this investigation—represents a generalisation of Tsallis entropy: While the first one is built from generator g_+, the second one is built from g. Further, the average of S_+ and S_- does assemble a third entropy:

$$S_0 = \sum_j^\Omega \frac{p_j^{-p_j} - p_j^{p_j}}{2} \ ,$$

which also belongs to the equivalence class $(c, d) = (1, 1)$ and rapidly converges to S_{BG}, as manifested by the following expansion

$$S_0 = -\sum_j^\Omega \left(p_j \ln p_j + \frac{1}{3!}(p_j \ln p_j)^3 + \frac{1}{5!}(p_j \ln p_j)^5 + \cdots \right) \ ,$$

although in this work we do not put special emphasis into this entropy, we shall reflect that the analytical features of S_+ and S_- are inherited to S_0.

2.4 Entropic Stability

The fact that an entropy satisfies the first three Shannon-Khinchin's axioms does not inevitably grant it the quality of representing physical systems. The latter is only assured if the entropy is stable[50]. That is, the entropy value will not change abruptly if the point where it has been evaluated is slightly shifted.

Namely, for a given $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that for two distributions of probability P and P', defined over the number of states Ω, it follows that $\|P - P'\|_1 < \delta_\epsilon$, hence $|S(P) - S(P')| < \epsilon S^*(\Omega)$, where S^* is the maximum of entropy. Therefore it is said that S is stable.

Lemma 1. *The non-extensive entropies S_+ and S_- are stable for any probability distribution P.*

Proof. Let $A_\pm(p;t)$ be defined as

$$A_\pm(P;t) = \sum_n^\Omega \left(p_j - e_\pm^{-t}\right) \theta\left(p_j - e_\pm^{-t}\right) \ , \tag{2.14}$$

where $\theta(x)$ is the Heaviside theta function and the stretched exponential functions $e_\pm$ are such that $\ln_\pm(e_\pm^x) = e_\pm^{\ln_\pm(x)} = x$, with $\ln_\pm(x) = (\pm 1 \mp x^{\pm x})/x$.

Since the inequality $|x\theta(x) - y\theta(y)| \leq |x - y|$ holds, it follows from (2.14) that $|A_\pm(P;t) - A_\pm(P';t)| \leq \|P - P'\|_1$. As well, the entropies $S_\pm$ are now expressed in terms of (2.14) as

$$S_\pm(P) = \phi + \int_0^\infty \mathrm{d}t \left[1 - A_\pm(P;t)\right] , \tag{2.15}$$

and we arrive at the following relations:

$$\begin{aligned} \left|S_\pm(P) - S_\pm(P')\right| &= \left| \int_0^\infty \mathrm{d}t [A_\pm(P;t) - A_\pm(P';t)] \right| \\ &\leq \int_0^{a_\pm + \ln_\pm \Omega} \mathrm{d}t \ \left|A_\pm(P;t) - A_\pm(P';t)\right| \\ &+ \int_{a_\pm + \ln_\pm \Omega}^\infty \mathrm{d}t \ \left|A_\pm(P;t) - A_\pm(P';t)\right| \ , \end{aligned} \tag{2.16}$$

where $a_\pm \geq -\ln_\pm \Omega$. In particular, if $a_\pm \geq 0$ we easily compute the integral in the first term, whereas the integral in the second term is performed by using $(1 - e_\pm^{-P+\ln_\pm \Omega}) \leq A_\pm(P;t) < 1$, then we obtain

$$\left|S_\pm(P) - S_\pm(P')\right| \leq \|P - P'\|_1(a_\pm + \ln_\pm \Omega) + e_\pm^{-a_\pm} R_\pm(\Omega)\,, \tag{2.17}$$

here $R_\pm(\Omega)$ are the residual functions from the integration of $e_\pm^{-t}$ w.r.t. t such that $R_\pm(\infty) \sim 1$. Moreover, given that $a_\pm \geq 0$, the right-hand side becomes minimum at $a_\pm = -\ln_\pm \|P - P'\|_1$, where $\|P - P\|_1 < 1$, therefore from (2.17) one gets

$$\frac{|S_\pm(P) - S_\pm(P')|}{\ln_\pm \Omega} \leq \delta\left(1 + \frac{R_\pm(\Omega)}{\ln_\pm \Omega}\right) - \delta \ln_\pm \delta\,, \tag{2.18}$$

this follows from the fact that $-x \ln_\pm x$ is a nonnegative, continuous function in the interval $[0, 1/e_\pm]$, in consequence $\|P - P'\|_1 < \delta < 1/e_\pm$ concluding that there is an appropriate δ_ϵ for every ϵ such that the right-hand side in (2.18) is a continuous function approaching 0 as $\delta \to 0$. And we conclude the proof. □

What Lemma 1 entails, is that S_+ and S_- enjoy the necessary attributes to truly represent physical systems. On top of that, given the arguments extracted from superstatistics, we suggest that any reasonable generalisation of (2.12) depending on $x \ln x$ should nearly coincide with the features already conceived in S_+ and S_- (or S_0).

2.5 Parametric Entropies

Although entropies S_+, S_- and S_{BG} depend only on the probability distribution, there are entropy functionals that, in addition, are subject to free parameters. Next, we survey some of the most representative cases usually found in the literature, to wit: Sharma-Mittal[67], Rényi[64] and Tsallis[75] entropies. We will see that unlike S_+ and S_-, parametric entropies do not have a well-defined thermodynamic limit.

2.5.1 Sharma-Mittal Entropy

This non-extensive entropy has the form $S = \phi\left(\sum_j g(p_j)\right)$, cf. Remark 1, and defines as:

$$S_{q,r} = \frac{1}{r-1}\left[1 - \left(\sum_j^\Omega p_j^q\right)^{\frac{1-r}{1-q}}\right], \qquad q, r \in \mathbb{R}\,, \tag{2.19}$$

by taking the limits $(q, r) \to 1$ and $r \to q$ we recover Boltzmann and Tsallis entropies, respectively.

Even $S_{q,r}$ does not possess the structure of (2.2), we are still entitled to classify it by the direct application of the asymptotic laws (2.9) and (2.10), yielding $(c, d) = (r, +\infty)$. In principle, due to the exponent d is unbounded, $S_{q,r}$ will satisfy SK1-SK3 whenever $0 < r < 1$, but the entropic stability will be dismissed.

To clarify this, consider an equipartition configuration, $p_j = 1/\Omega$, then $S_{q,r}$ becomes

$$S_{q,r} = \frac{\Omega^{1-r} - 1}{1 - r} = \frac{\exp[(1-r)S_{BG}] - 1}{1 - r} \,, \tag{2.20}$$

in the limit $r \to 1$ we recover the standard expression given by Boltzmann, otherwise the entropy will exponentially separate from S_{BG} as Ω increases, no matter how close r is from the unit, as we shall prove now.

First, we solve Ω from (2.20) and use $S_{BG} = \ln \Omega$ to express the Boltzmann-Gibbs entropy in terms of $S_{q,r}$ as

$$S_{BG} = \lim_{r \to 1} \frac{1}{r - 1} \ln[1 + (1 - r)S_{q,r}] \,, \tag{2.21}$$

expanding in series, we get

$$S_{BG} = \lim_{r \to 1} \left[S_{q,r} + \frac{1}{2}(r-1)S_{q,r}^2 + \frac{1}{3}(r-1)^2 S_{q,r}^3 + \frac{1}{4}(r-1)^3 S_{q,r}^4 + \cdots \right] \,,$$

in the limit $r \to 1 + \epsilon$, for ϵ arbitrarily small, the former becomes

$$S_{BG} \neq S_{q,1+\epsilon} + \frac{1}{2}\epsilon S_{q,1+\epsilon}^2 + \frac{1}{3}\epsilon^2 S_{q,1+\epsilon}^3 + \frac{1}{4}\epsilon^3 S_{q,1+\epsilon}^4 + \cdots \,,$$

however, as seen from (2.20), $S_{q,1+\epsilon}$ will diverge for $\epsilon \ll 1$. Therefore, $S_{q,r}$ will never reach the thermodynamic limit, regardless of the value of ϵ.

2.5.2 Rényi Entropy

In turn, we shall expand (2.19) in powers of r in the vicinity of the unit sphere:

$$S_{\alpha,r} = \frac{\ln x}{\alpha - 1} + \frac{(r-1)\ln^2 x}{2(\alpha-1)^2} + \frac{(r-1)^2 \ln^3 x}{6(\alpha-1)^3} + \frac{(r-1)^3 \ln^4 x}{24(\alpha-1)^4} + \cdots \,,$$

where $x = \sum_j^\Omega p_j^\alpha$. Taking the limit $r \to 1$ and substituting x, we have the Rényi entropy:

$$S_\alpha = \frac{\ln\left(\sum_j^\Omega p_j^\alpha\right)}{1 - \alpha} \,, \tag{2.22}$$

where α is a parameter associated with the degree of convexity; actually those values outside $0 < \alpha \leq 1$ could compromise the axioms SK2-SK3. Nonetheless, values $\alpha > 1$ have a place in the study of magnetic anisotropies[28].

This entropy is additive and belongs to the equivalence class $(c, d) = (1, 1)$, which is rapidly verified by direct application of (2.9) and (2.10) or, alternatively, by maximising S_α subject to the normalisation constraint $\sum p_j = 1$, hence obtaining $S_{BG} = S_\alpha = \ln \Omega$.

2.5.3 Tsallis Entropy

As for this entropy, it can be derived in several ways—for the initial proposal see Ref.[37]. One of them, as aforementioned, results from the substitution $r \to q$ in Eq. (2.19). Interestingly, we also obtain this entropy from (2.22) by the following expansion:

$$\begin{aligned} S_\alpha &= -\frac{\ln\left(\sum_j^\Omega p_j^\alpha\right)}{r-1} = -\frac{\sum_j^\Omega p_j^\alpha - 1}{r-1} + \frac{(\sum_j^\Omega p_j^\alpha - 1)^2}{2(\alpha-1)} + \cdots \\ &\underset{\alpha\to q}{\approx} \frac{1 - \sum_j^\Omega p_j^q}{q-1} = S_q \,, \end{aligned} \tag{2.23}$$

thus yielding the Tsallis entropy iff $|\sum p_n^q - 1| \ll 1$. Much alike the case of Rényi, the free parameter q should be restricted to values in $(0, 1)$ to avoid violations to SK2 or SK3. Nevertheless, in some studies on turbulence[5], apparently consistent outcomes have been achieved taking into account values $q > 1$.

Tsallis entropy belongs to the equivalence class $(c, d) = (q, 0)$. In the limit $q \to 1$ it resembles S_{BG} and, in general, there is no well-defined thermodynamic limit for this entropy, as is evident from the following argument. If we maximise S_q subject to the normalisation condition $\sum p_j = 1$, its maximum value is reached at $p_j = 1/\Omega$ for all j. Therefore, for a microcanonical ensemble, we have:

$$S_q = \frac{\Omega^{1-q} - 1}{1-q} = \frac{\exp[(1-q)S_B] - 1}{1-q} \,, \tag{2.24}$$

which is nothing but Eq. (2.20). It follows that, by the same reasons already exposed in Sec. 2.5.1, S_q will diverge from S_{BG} as the number of available states grows, see Figure 2.3.

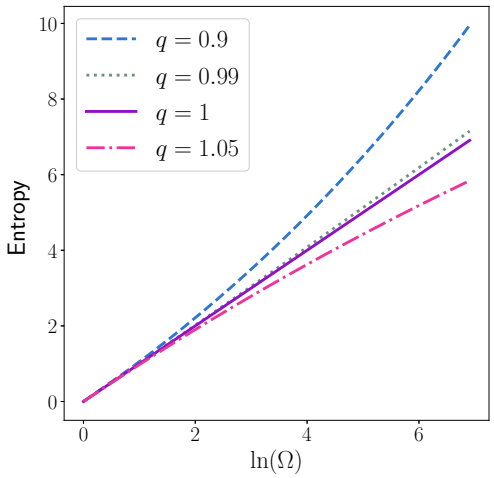


Figure 2.3: Tsallis entropy - The behaviour of Tsallis entropy for different values of q according to Eq. (2.24). Compare with the thermodynamic compatibility of S_+ and S_- in Fig. 2.1.

3

Generalised Stochastic Dynamics

In this chapter, we focus on an entropy approach to obtain generalised Fokker-Planck equations. We are particularly interested in those Fokker-Planck equations derivable from the non-extensive entropies S_+ and S_- introduced in Chapter 2. As a result, the generalised models will include nonlinear terms, which are interpreted as corrections to the usual mean-field Fokker-Planck equation.

As we have shown in Chapter 2, S_+ and S_-, as well as S_{BG}, belong to the same equivalence class, suggesting they are thermodynamically compatible. This is a pleasing result, in that the generalised Fokker-Planck equations leveraged by S_+ and S_- will bring suitable corrections to the mean-field Fokker-Planck equation in a scenario where the accessible microstates are comparable to the number of particles.

Besides, given that S_+ and S_- were originally obtained from the superstatistics framework, the resulting Fokker-Planck equations will inherit the nonequilibrium attributes bestowed to these entropies. We shall discuss the anomalous diffusion the generalised Fokker-Planck equations exhibit, while accompanying our exposition with some numerical experiments. We remark that there are diffusion models in biological sciences, whose structure coincides with that of our models.

3.1 Asymptotic Scaling of Dynamical Distributions

The asymptotic classification introduced in Chapter 2 consists of two scale exponents (c, d) granted by Eqs. (2.9) and (2.10). These exponents form an equivalence class that groups entropies into a certain category. However, in case that we are interested in time-

dependent problems, a third exponent shall be introduced to observe the asymptotic scaling of time-dependent probability distributions. Those distributions are customarily studied in stochastic processes, e.g. the transport phenomena described by the Fokker-Planck equation[27,61].

Exploratory generalisations of the Fokker-Planck equations and the corresponding asymptotic scalings were calculated from some non-extensive entropies in Refs.[19,23]; in our opinion a noteworthy drawback enters the picture as those non-extensive entropies dismiss the stability criterion, cf. Sec. 2.4. In contrast, our purpose in this chapter is to derive generalised Fokker-Planck equations associated with S_+ and S_-, which are stable entropies. To this aim, we need in the first place examining the time-dependent scaling laws.

Some of these attempts come from the phenomenological classification of transport phenomena[25] following the rescaling of a given distribution $\rho = \rho(\mathbf{x}, t)$ by a factor of the type $\tau^{-\gamma}$. In this respect, the distribution $\rho(\mathbf{x}, t)$ subjects to the condition that

$$\rho(\mathbf{x}, t) = \tau^{-\gamma} \rho\left(\frac{\mathbf{x}}{\tau^{\gamma}}, \frac{t}{\tau}\right) \tag{3.1}$$

remains invariant under rescaling of space $\mathbf{x}$ and time t coordinates[7,53].

In turn, let $\mathbf{J}(\mathbf{x}, t)$ be the flux (or probability current), then by condition (3.1) the following continuity equation must be satisfied:

$$\tau^{-(\gamma+1)} \partial_t \rho = -\tau^{-\gamma(c+2)} \nabla \cdot \mathbf{J}(\mathbf{x}, t) \ , \tag{3.2}$$

given that equality must hold, we get $\gamma = 1/(c+1)$, which suggests the role exponent c has in the classification of dynamical systems.

In this context, now the exponents (γ, c, d) amount to an equivalence class for time-dependent systems. Accordingly, the equivalence class $(\gamma, c, d) = \left(\frac{1}{2}, 1, 1\right)$, which relates to S_+, S_- and S_{BG}, suggests the three entropies enjoy the same asymptotic time-scale attributes, while S_+ and S_- will modestly separate from S_{BG} at early times of evolution of a stochastic system.

3.2 Generalised Fokker-Planck Equations

Analogously to the Langevin equation in Brownian motion, the Fokker-Planck equation models the time evolution problem of a density function associated with the movement

of a particle governed by stochastic and drifting forces. Depending on the application, the density may be related to distinct observables, such as the velocity or the position of a particle.

There are various generalisations of the Fokker-Planck equation, e.g. both the Kirkwood[46] and Rice-Allnatt[65] theories of transport phenomena in dense fluids, derive Chandrasekhar equations[17] as generalisations of the Fokker-Planck equation to phase space. Listing all the generalisations, however, would be challenging and perhaps a fruitless endeavour. Thus, we shall limit our discussion to those generalisations in the configuration space, having the form:

$$\partial_t \rho = \nabla \cdot [D(\rho)\nabla F[\rho] + \chi(\rho)\rho\nabla\Phi] \ , \tag{3.3}$$

where $D(\rho)$ and $\chi(\rho)$ are the diffusion and effective drift coefficients, $\Phi(\mathbf{x})$ is the potential field wherein Brownian particles are assumed to move and $F[\rho]$ is a Lyapunov functional oftentimes referred as effective density. As we will see, the functional dependence of $F[\rho]$ and $\chi(\rho)$ shall be determined from the measure of entropy with the possibility of involving nonlinearities in ρ.

Instead the Fokker-Planck equation is also expressed in a conservative form by associating Eq. (3.2) with Eq. (3.3), hence, we identify:

$$\mathbf{J}(\mathbf{x},t) = -\left(D(\rho)\nabla F[\rho] + \chi(\rho)\rho\nabla\Phi\right) \ , \tag{3.4}$$

on the right-hand member, the first and the second terms convey the diffusion and drifting responses, respectively. Accordingly, Eq. (3.3) becomes

$$\begin{aligned} \partial_t \rho &= -\nabla \cdot \mathbf{J} \\ &= -\nabla \cdot (\mathbf{J}_{\text{diff}} + \mathbf{J}_{\text{drift}}) \ . \end{aligned} \tag{3.5}$$

As aforementioned, the measure of entropy is the route to obtain the functionals $F[\rho]$ and $\chi(\rho)$. In the case of the effective density $F[\rho]$, it is leveraged from the notion of generalised logarithms[36] $\Lambda(\rho)$, whilst the relation between the diffusion coefficient and the entropic form leads to the generalised drift coefficient. In turn, we shall introduce their formal definitions.

Definition 5. *Let $g(\rho)$ be the entropic form, as defined in (2.2), and let $\omega(\rho)$ be the energy weight, such that*

$$w(\rho) = \left(1 + \frac{\alpha}{E^*}\right) \int_0^\rho \frac{\mathrm{d}x}{1 - E(x)/E^*} \ , \tag{3.6}$$

where E is the inverse of the Boltzmann factor (2.4) whose minimum value is E^* (if any) and $\alpha = -\int_0^1 \mathrm{d}y E(y)$, then the generalised logarithm is defined as:

$$\Lambda(\rho) = \frac{g'(\rho) - \alpha}{\beta w'(\rho)} , \quad \Lambda(1) = 0 , \quad \Lambda'(1) = 1 , \tag{3.7}$$

which results from the maximisation of the functional:

$$\Gamma = g(\rho) - \alpha\rho - \beta w(\rho) E . \tag{3.8}$$

Expressly, the effective logarithm results from the maximisation of the entropic form subject to conditions of normalisation of the density and conservation of energy, with α and β as Lagrange multipliers.

Definition 6. *Let $\Lambda(\rho)$ be a generalised logarithm, we formally define the effective density as*[23]*:*

$$F[\rho] = -\beta \int_0^\rho \mathrm{d}x \, x \, \partial_x \Lambda(x) . \tag{3.9}$$

Remark 5. *The functional in (3.9) may be nonlinear in ρ and owns the required information to ponder the changes in diffusion throughout space. The particular case $F[\rho] = \rho$, resembled by S_{BG}, corresponds to the mean field Fokker-Planck equation.*

Definition 7. *Let $D(\rho)$ be the diffusion coefficient and $\chi(\rho)$ the effective drift coefficient. From Eq. (3.3), we introduce the notations*

$$D\zeta(\rho) = \frac{\mathrm{d}}{\mathrm{d}\rho}\rho D(\rho), \quad \chi\xi(\rho) = \chi(\rho)\rho , \tag{3.10}$$

then the effective drift weight fulfils the relation[19]*:*

$$\xi(\rho) = -\frac{\zeta}{g''(\rho)} \geq 0 , \tag{3.11}$$

such that to preserve the Einstein relation, the temperature is defined by

$$T = \frac{D}{\chi} . \tag{3.12}$$

To illustrate how this formalism works, let us consider the Boltzmann-Gibbs entropy. In this case the Boltzmann factor is $B(E) = x = \exp(-\beta E)$, then $E(x) = -\frac{1}{\beta}\ln x$. It follows from Eq. (3.6) that the energy must be weighted by $w(\rho) = \rho$ and the effective logarithm (3.7) is $\Lambda_{BG}(\rho) = -\frac{1}{\beta}\ln x$. Substituting the effective logarithm into (3.9), it yields

$$F[\rho] = \int_0^\rho \mathrm{d}x \, x \frac{1}{x} = \rho .$$

As for (3.11) and assuming the diffusion coefficient is constant, we have

$$\xi(\rho) = \frac{\zeta}{1/\rho} \propto \rho \,,$$

therefore, the resulting mean-field Fokker-Planck equation is

$$\partial_t \rho = D\nabla \cdot \left[\nabla\rho + \frac{1}{T}\rho\nabla\Phi_{\text{ext}} \right] \,, \tag{3.13}$$

this equation is commonly referred as the Smoluchowski equation, describing the Brownian motion of classical particles in a fixed potential field. For example, if Φ_{ext} is the gravitational potential, we would have a model for self-gravitating Brownian particles, that is:

$$\begin{aligned} \partial_t \rho &= D\nabla \cdot \left[\nabla\rho + \frac{1}{T}\rho\nabla\Phi_{\text{ext}} \right] \,, \\ \nabla^2\Phi_{\text{ext}} &= S_d G\rho \,, \end{aligned}$$

G is the universal gravitational constant and $S_d = 2\pi^{\frac{d}{2}}/\Gamma\left(\frac{d}{2}\right)$ is the solid angle in d dimensions. Likewise, the Poisson equation can be generalised by measures of entropy, we shall not enter into such discussion however.

Further models are obtained from Eq. (3.3) by selecting a particular measure of entropy and turning on or off the potential field. We list a few of them in Table 3.1.

Model	$D(\rho)$	$F[\rho]$	$\chi\xi(\rho)$	Φ	Entropy
Smoluchowski	D	ρ	$\chi\rho$	Φ_{ext}	S_{BG}
Debye & Hückel	D	ρ	$\chi\rho$	Φ_{electric}	S_{BG}
Diffusion	D	ρ	0	0	S_{BG}
Porous medium	D	ρ^q	0	0	S_q
Plastino	D	ρ^q	$\chi\rho^q$	Φ_{ext}	S_q

Table 3.1: Some models characterised through the Fokker-Planck equation (3.3).

3.3 Models with Asymptotic-Boltzmann Behaviour

In this section, we are to obtain the Fokker-Planck equations that correspond to the non-extensive entropies S_+ and S_-. We point up these non-extensive entropies are stable for any probability distribution—regardless of the number of available states—in contrast with parametric entropies such as S_q and S_α, which exhibit instabilities and therefore cannot represent experimentally observable quantities. Those Fokker-Planck equations derived from unstable entropies will undoubtedly inherit the same unfavourable etiquette, and the dynamics they describe shall be regarded as devoid of physical meaning.

The differential equations that we are to obtain are nonlinear in ρ, nonetheless the linear response is refurbished once the process involved attains the equilibrium or whenever the number of available states is much greater than the number of particles playing a role.

To begin with, we recall the effective Boltzmann factors (2.7) and (2.8) univocally linked to S_+ and S_-:

$$B_\rho^+(E) = (1+\rho\beta_0 E)^{-\frac{1}{\rho}} , \quad B_\rho^-(E) = (1-\rho\beta_0 E)^{+\frac{1}{\rho}} ,$$

solving for E in both cases and substituting into (3.6), we obtain the corresponding energy weights:

$$w_+(\rho) = \rho^{\rho+1} , \quad w_-(\rho) = \rho^{\rho-1} .$$

Given the weights w_+ and w_-, we readily compute the effective logarithms

$$\Lambda_+(\rho) = \frac{1}{\beta}\frac{1-\rho^{-\rho}+\ln\rho}{1+\rho+\rho\ln\rho}, \tag{3.14}$$

and

$$\Lambda_-(\rho) = \frac{1}{\beta}\frac{1-\rho^{\rho}+\ln\rho}{1-\rho-\rho\ln\rho}. \tag{3.15}$$

These logarithms converge to Λ_{BG} for low and high ρ, i.e. they show an equivalent asymptotic behaviour as the one already reviewed for S_+ and S_-, see Fig. 3.1. Hence we anticipate a smooth transition, from our models to the dynamics described by the mean-field Fokker-Planck equation, as the process approximates to the equilibrium.

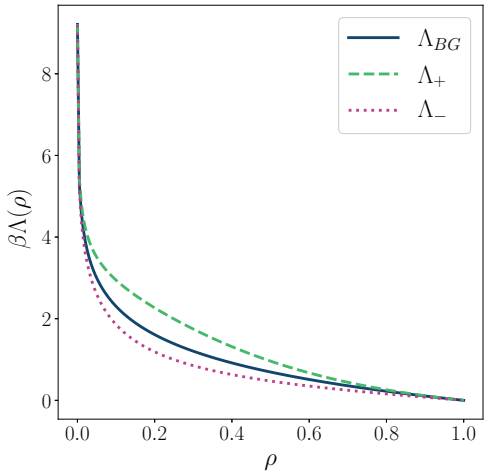


Figure 3.1: Effective logarithms related to S_+, S_- and S_{BG} - The three effective logarithms tend to converge for low and high densities, whereas smooth differences manifest at in-between regimes.

To proceed, we now compute the effective densities $F_+[\rho]$ and $F_-[\rho]$ by substituting Eqs. (3.14) and (3.15) into (3.9), we obtain

$$\begin{aligned} F_+[\rho] &= \int_0^\rho \mathrm{d}x \left(\frac{1 + x^{1-x} + x^{1-x}\ln x}{1 + x + x\ln x} - \frac{(\ln x + 2)\left(x - x^{1-x} + x\ln x\right)}{(1 + x + x\ln x)^2} \right) \\ &= \rho + \frac{\rho^2}{4} + \frac{\rho^3}{27} + \frac{\rho^4}{128} + \cdots , \end{aligned} \tag{3.16}$$

analogously, we have

$$\begin{aligned} F_-[\rho] &= \int_0^\rho \mathrm{d}x \left(\frac{1 - x^{1+x} - x^{1+x}\ln x}{1 - x - x\ln x} + \frac{(\ln x + 2)\left(x - x^{1+x} + x\ln x\right)}{(1 + x + x\ln x)^2} \right) \\ &= \rho - \frac{\rho^2}{4} + \frac{\rho^3}{27} - \frac{\rho^4}{128} + \cdots , \end{aligned} \tag{3.17}$$

as expected, the nonlinear terms are monotonically subdominant provided the density ρ is normalised and the accompanying coefficients diminish progressively.

The effective drift weight must be obtained from Eq. (3.11), to this aim we assume the diffusion coefficient is constant, then we obtain:

$$\xi_+(\rho) = \frac{\rho^{1-\rho}}{1 + (1 + \ln\rho)^2\rho} , \quad \xi_-(\rho) = \frac{\rho^{1+\rho}}{1 - (1 + \ln\rho)^2\rho} , \tag{3.18}$$

in the limiting case $\rho \ll 1$ both expressions resemble $\xi_{BG}(\rho) = \rho$, as presumed, yet for high densities we shall stick to (3.18). Note the weight ξ_- blows up as the density approaches the unit, this flaw shall be removed by introducing a cutoff condition, which we will discuss later in our numerical computations.

In summary, from Eqs. (3.16)-(3.18) and (3.3) the generalised Fokker-Planck equations asymptotically equivalent to the mean-field Fokker-Planck equation (3.13) are

$$\partial_t \rho = -\nabla \cdot \mathbf{J}_+ \ , \quad \partial_t \rho = -\nabla \cdot \mathbf{J}_- \ , \tag{3.19}$$

where

$$\mathbf{J}_+(\mathbf{x},t) = -D\left[\nabla\left(\rho + \frac{\rho^2}{4} + \frac{\rho^3}{27} + \cdots\right) + \frac{1}{T}\frac{\rho^{1-\rho}}{1+(1+\ln\rho)^2\rho}\nabla\Phi_{\text{ext}}\right] \ , \tag{3.20}$$

and

$$\mathbf{J}_-(\mathbf{x},t) = -D\left[\nabla\left(\rho - \frac{\rho^2}{4} + \frac{\rho^3}{27} - \cdots\right) + \frac{1}{T}\frac{\rho^{1+\rho}}{1-(1+\ln\rho)^2\rho}\nabla\Phi_{\text{ext}}\right] \ , \tag{3.21}$$

for numerical purposes, we shall eventually truncate the diffusion term $\mathbf{J}_{\text{diff}}$ (first term inside brackets) to third-order in both cases, higher-order terms indeed will barely contribute to the dynamical picture.

Besides, Eqs. (3.19) suggest is the mass distribution $M = \int \mathrm{d}^3x\, \rho$ is a conserved quantity, then the normal component of each $\mathbf{J}_\pm$ will vanish at the boundary, which turns evident by applying the divergence theorem to (3.19):

$$\partial_t \int_V \mathrm{d}^3x\, \rho = -\oint_S \mathrm{d}\mathbf{S} \cdot \mathbf{J}_\pm \ . \tag{3.22}$$

Contrary, whenever a system is thermodynamically open it will experience the exchange of energy and matter with the environment, then the analysis of solutions can be put in terms of a free energy with the form of a Lyapunov functional[18,19].

We cannot give closed solutions to the generalised Fokker-Planck equations (3.19), except when the potential is neglected, in which case only semi-analytical solutions are stated—as we discuss next. We will also present some numerical solutions for some idealised physical models in Sec. 3.4.

3.3.1 Anomalous Diffusion

Next, we shall neglect the drift terms in the Fokker-Planck equations (3.19), staying only with the generalised components of diffusion, that is

$$\partial_t \rho = D\nabla^2 \left(\rho + \frac{\rho^2}{4} + \frac{\rho^3}{27} + \cdots \right) , \quad \partial_t \rho = D\nabla^2 \left(\rho - \frac{\rho^2}{4} + \frac{\rho^3}{27} - \cdots \right) . \tag{3.23}$$

An interesting feature these nonlinear diffusion equations possess is the appearance of an effective drift term conveyed by the carriage of nonlinearities. To illustrate this fact, observe the effective densities (3.16) and (3.17) are alternatively expressed in the form

$$F_\pm[\rho] = \int_0^\rho \mathrm{d}x \, R_\pm(x) , \tag{3.24}$$

hence the generalised currents become:

$$\mathbf{J}_\pm(\rho) = -D\nabla \left[\int_0^\rho \mathrm{d}x \, R_\pm(x) \right] , \tag{3.25}$$

these currents are each one proportional to the chemical potential[21]:

$$\mathbf{J}_\pm = -D\nabla \mu_\pm(\rho) , \tag{3.26}$$

even more, provided the chemical potential $\mu(\rho)$ comes from the gradient of a function $f[\rho]$, namely $\mu(\rho) = f'[\rho]$, then by association we identify

$$f'_\pm[\rho] = \int_0^\rho \mathrm{d}x \, R_\pm(x) , \tag{3.27}$$

and finally, from the continuity equation, we obtain:

$$\partial_t \rho = \nabla \cdot [D R_\pm(\rho) \nabla \rho] , \tag{3.28}$$

where we recognise the generalised diffusion coefficients of our models:

$$D_\pm(\rho) = D R_\pm(\rho) , \tag{3.29}$$

of course, the standard diffusion coefficient is reconstructed at first-order approximation from either the effective density in Eq. (3.16) or (3.17). Likewise, the nonlinear diffusion equation (3.28) are expressed in terms of $f_\pm[\rho]$ as

$$\partial_t \rho = D\nabla \cdot [\partial_{\rho\rho} f_\pm[\rho] \nabla \rho] , \tag{3.30}$$

where

$$f_{\pm}[\rho] = \frac{\rho^2}{2} \pm \frac{\rho^3}{12} + \frac{\rho^4}{108} \pm \cdots \,, \tag{3.31}$$

therefore Eq. (3.30) becomes:

$$\partial_t \rho = D\left[\nabla^2 \rho + \varphi(\rho)\nabla\rho\right], \tag{3.32}$$

where $\varphi(\rho)$ contains all the ρ-dependent terms in $f''[\rho]$.

We notice that Eq. (3.32) induces an effective drift term weighted by $\varphi(\rho)$—which can be connected to congregation-diffusion models[77]—while preserving the ordinary diffusion term. In summary, the contribution of the nonlinearities in Eq. (3.23) stands for effective drifting forces in a regime of high densities, otherwise, at low densities the ordinary diffusion will be recovered without picking up the presence of the effective drift terms.

3.4 Numerical Solutions

We are right away to look at some numerical solutions to the nonlinear Fokker-Planck equations (3.19) considering specific external potentials $\Phi = \Phi_{\text{ext}}$. As it was discussed in Sec. 3.3.1, the nonlinearities in the diffusion terms of $\mathbf{J}_{\pm}$ convey an effective drift behaviour that we shall test numerically. Additionally, we compare our generalised diffusion models with the chemotaxis-aggregation approach studied by P. Turchin[77]. To end this discussion, we introduce a transient-diffusion modification into our models, which agrees with experiment and the respective phenomenological model whenever long relaxation times are appointed in our non-equilibrium approach.

3.4.1 Effective Equilibrium

We have numerically integrated Eqs. (3.13) and (3.19) (truncating $F_{\pm}$ to third order) in 1+1 dimensions, see Fig. 3.2. In the figure it is shown the time evolution of a Gaussian distribution in the presence of an attractive linear potential $\Phi = -5x$ in the domain $x \in [-20, 20]$ and evolving along the interval $t \in [1, 10]$, according to the boundary and initial conditions:

$$\rho(x = -20, t) = \rho(x = 20, t) = \frac{e^{-\frac{100}{t}}}{2\sqrt{\pi t}} \,,$$

and

$$\rho(x, t = 1) = \frac{e^{-\frac{x^2}{4}}}{2\sqrt{\pi}} ,$$

in our numerical example we have considered the diffusion coefficient $D = 1$.

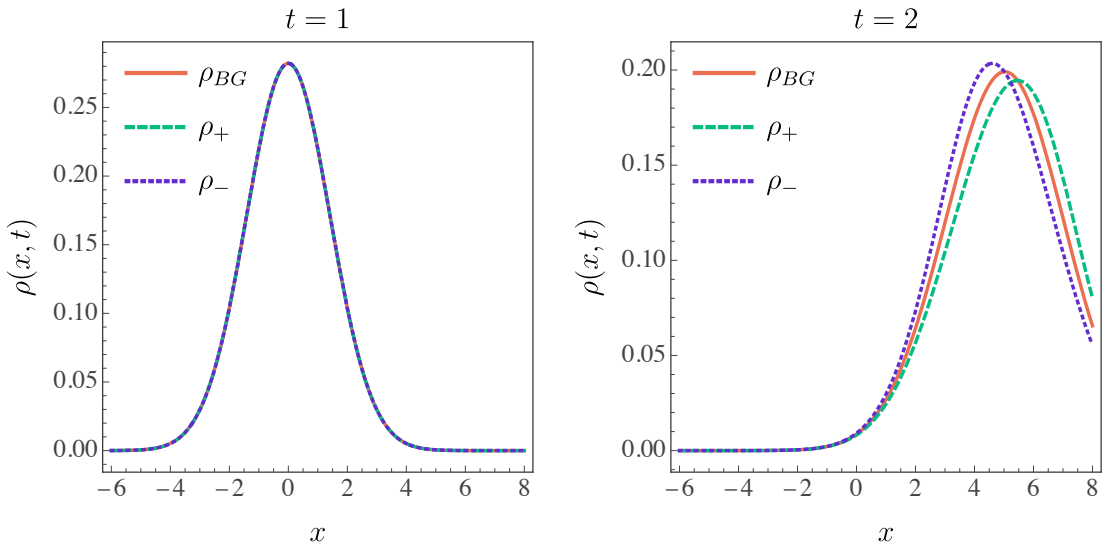


Figure 3.2: Solutions to the generalised Fokker-Planck equations - The drifting behaviour at different times of ρ_+, ρ_- and ρ_{BG}, which are solutions to the Fokker-Planck equations (3.13) and (3.19), respectively, in the presence of a linear external potential. The density ρ_+ drifts faster than the other solutions, i.e. the system is more sensitive to the response of the linear potential.

The example in Fig. 3.2 confirms the solution ρ_+ evolves faster than the standard solution ρ_{BG} and ρ_-. This latter case becomes resistant to the influence of the attractive linear potential and spreads slower than the other two cases.

To enlarge our discussion next we shall examine the time evolution of Gaussian distributions in 2+1 dimensions. In our following numerical experiment, we considered diffusion coefficients $D_x = 5$ and $D_y = 1/3$ along the Ox and Oy directions, and the external potential $\Phi(x, y) = -x - y$ with a drift coefficient $\chi = 1$. We integrated Eqs. (3.13) and (3.19) in a square domain $x, y \in [-10, 10]$ within the time interval $t \in [1/2, 15]$, and set the boundary and initial conditions:

$$\rho(x = 10, y, t) = \rho(x = -10, y, t) = \frac{e^{-\frac{y^2}{4t} - \frac{25}{4t}}}{4\pi t} ,$$

and

$$\rho(x, y = 10, t) = \rho(x, y = -10, t) = \frac{e^{-\frac{x^2}{4t} - \frac{25}{4t}}}{4\pi t} ,$$

with

$$\rho\left(x, y, t = \frac{1}{2}\right) = \frac{e^{-\frac{x^2}{2} - \frac{y^2}{2}}}{2\pi} ,$$

the results are shown in Fig. 3.3, where each row corresponds to the solutions $\rho_{BG}, \rho_+, \rho_-$, in that order, and each column relates to the evolution times $t = \frac{1}{2}, t = 1, t = \frac{3}{2}, t = 2$. Observe how the density ρ_- propagates reticently in comparison with the other densities, while ρ_+ exhibits the fastest diffusion of the three cases.

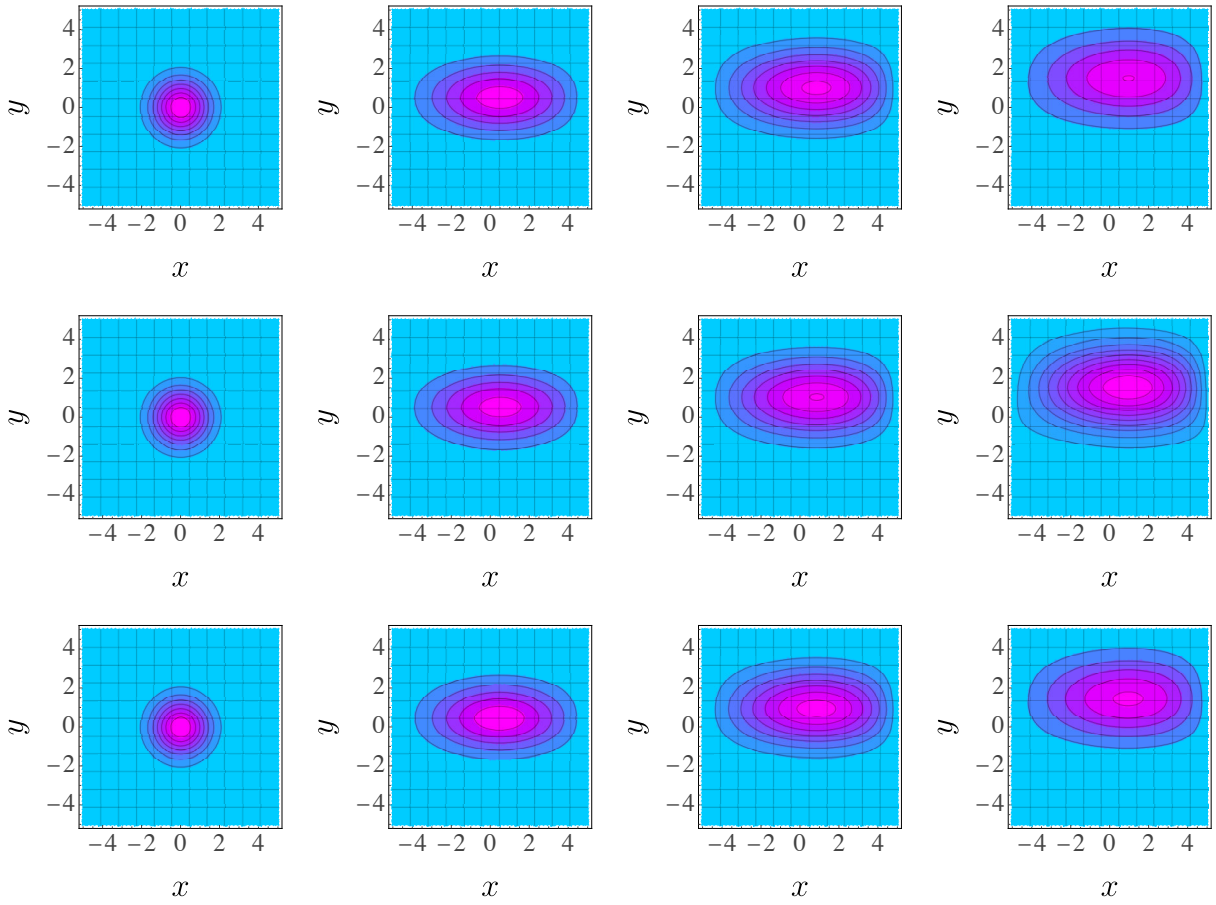


Figure 3.3: Solutions to the Fokker-Planck equations in 2+1 dimensions - Each row from top to bottom corresponds to the solutions ρ_{BG}, ρ_+ and ρ_-, respectively, to the generalised Fokker-Planck equations (3.13) and (3.19) in 2+1 dimensions at evolution times $t = \frac{1}{2}, t = 1, t = \frac{3}{2}, t = 2$, from left to right, responding to the potential $\Phi = -x - y$. Observe how the density ρ_- evolves slower than the other solutions.

As another example, we now consider the same boundary and initial conditions as for the previous numerical experiment, but this time implementing the potential $\Phi(x,y) = -\text{sech}\left(\frac{x}{2}-2\right) - \text{sech}\frac{y}{2}$ and the diffusion coefficients $D_x = 2$ and $D_y = 8$, while the drift coefficient is fixed with $\chi = 1$. The time evolution of the densities ρ_{BG}, ρ_+ and ρ_- is displayed in Fig. 3.4. The spread of ρ_- (third row) manifests a slower diffusion and drift behaviour in contrast with ρ_+ (second row), which reacts faster than the other densities in the presence of the same potential.

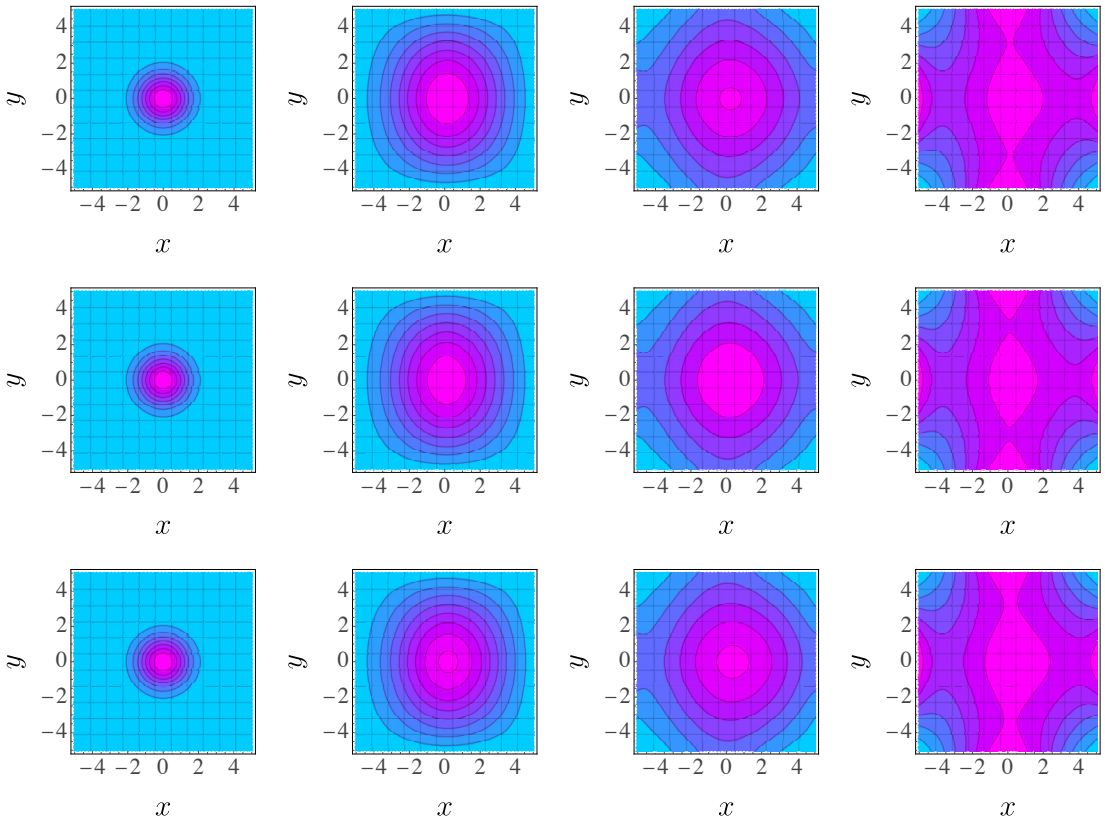


Figure 3.4: Time Evolution of the solutions to the Fokker-Planck equations - Similar to the example in Fig. 3.2, each row from top to bottom corresponds to the solutions ρ_{BG}, ρ_+ and ρ_-, respectively, to Eqs. (3.13) and (3.19) in 2+1 dimensions at evolution times $t = \frac{1}{2}, t = 2, t = 3, t = 4$, from left to right, responding to the potential $\Phi = -\text{sech}\left(\frac{x}{2}-2\right) - \text{sech}\frac{y}{2}$.

3.4.2 Chemotaxis Aggregation Models

In this section we look at a biological application to describe the random behaviour of bacterial populations or aggregation movements in ecology, which are modelled by the chemotaxis-aggregation diffusion equation proposed in Ref.[77]. This model comprises random walks of length λ with average times τ and a maximum degree of attraction bias k_0, namely:

$$\begin{aligned} \partial_t \rho &= \nabla^2 \left[D\rho - \kappa\rho^2 + \frac{2\kappa}{3\omega}\rho^3 \right] \\ &\equiv \nabla^2 F_{\text{agg}} \ , \end{aligned} \tag{3.33}$$

where $\kappa = k_0\lambda^2/\tau$ and ω stands for the critical density that turns the movement from attractive into repulsive. We are to compare this equation to the generalised diffusion equations (3.23) or (3.30). The similitudes between F_{agg} and $F_\pm$ are direct: The three effective densities possess the same structure if $F_\pm$ are to third-order approximated, except F_{agg} enjoys of free parameters (κ, ω) that are phenomenologically adjusted.

The parameter κ in Eq. (3.33), indicates either of two types of movement: if $\kappa > 0$, then there is a tendency to move away from conspecifics, otherwise there is a tendency to move towards conspecifics. Comparing F_{agg} with F_+ and F_-, this means that the effective density F_+ characterises an attractive movement between conspecifics for a fixed critical density $\omega < 0$, whereas F_- describes a repulsive movement for $\omega > 0$. In addition, notice the nonlinear terms in F_+ are always positive, thus concentrating high densities, whereas the nonlinear terms in F_- represent low concentration densities. In practice, however, organisms usually aggregate at low densities.

Furthermore, our equations (3.23) exhibit nonlinear diffusion for regions of space where the system experiences a sort of faint interactions before reaching the equilibrium, in other words, those regions where the confined constituents are arranged such that the resulting interaction forces are not entirely negligible.

The behaviour manifested by each of the effective densities F_+, F_- and F_{agg} is shown in Fig. 3.5. As for the aggregation model, it will account for the diffusion driven by a congregation movement mode inasmuch as the dynamic level is above that of the usual diffusion[77]. This behaviour, however, will tend to transform as the density increases, which means the movement will retract until its character finally swaps, thus turning into a weakly congregation mode. The inflection points are regarded as equilibrium points.

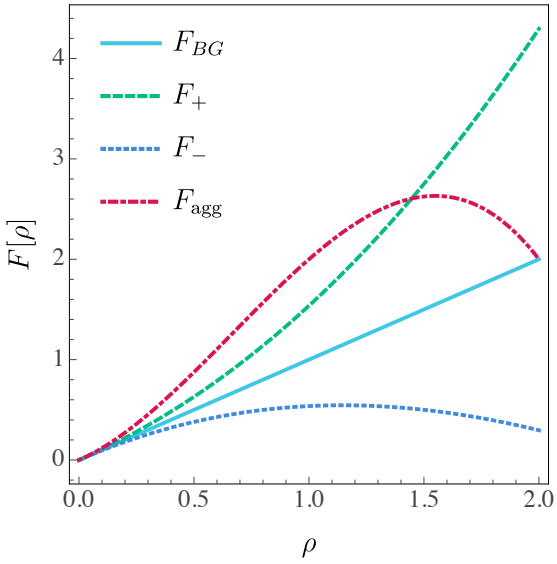


Figure 3.5: Diffusion levels - Diffusion levels characterised by the effective densities F_+, F_-, F_{agg} compared with the ordinary level $F_{BG} = \rho$. In the case of F_{agg}, we fixed the parameters $\kappa = -2$ and $\omega = 8/6$. Those densities lying in the region above (below) ρ will spread faster (slower) than the usual linear diffusion.

The chemotaxis-aggregation models are worth a broad application in biology and ecology[54]. In those contexts, the nonlinear terms appearing in the effective density functions convey effective fluctuating forces resulting from either the interaction between bacterial populations and the concentration of acrasin—in the case of biology—or between animal species and their surroundings—in the case of ecology. In both cases, the phenomenon consists of collective movements throughout a region where environmental conditions will favour or prevent such transit. However, in the case of our models, the effect of interactions will vanish as the number of available states becomes much grater than the number of individuals.

As a numerical example, we have integrated Eqs. (3.23) and (3.33) in 1+1 dimensions, in the spatial domain $x \in [-20, 20]$ along the evolution interval $t \in [1/2, 10]$, with the following boundary and initial conditions:

$$\rho(x = -20, t) = \rho(x = 20, t) = \frac{e^{-\frac{100}{t}}}{2\sqrt{\pi t}} \, ,$$

and

$$\rho\left(x, t = \frac{1}{2}\right) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \, ,$$

the outcomes are displayed in Fig. 3.6. Observe how ρ_{agg} experiences a retarded diffusion response in contrast with ρ_+ and ρ_-. The reason is rather for this simulation we set $\kappa = 3$, which is the coefficient accompanying the nonlinear term $-\rho^2$, thus exerting a stronger effective opposition to spreading than the other two cases.

In contrast, see Fig. 3.7, where we have integrated Eqs. (3.23) and (3.33) exactly the same way as in the previous example, but swapping the sign of κ. Thus ρ_{agg} describes a type of congregation slightly stronger than the one estimated by ρ_+.

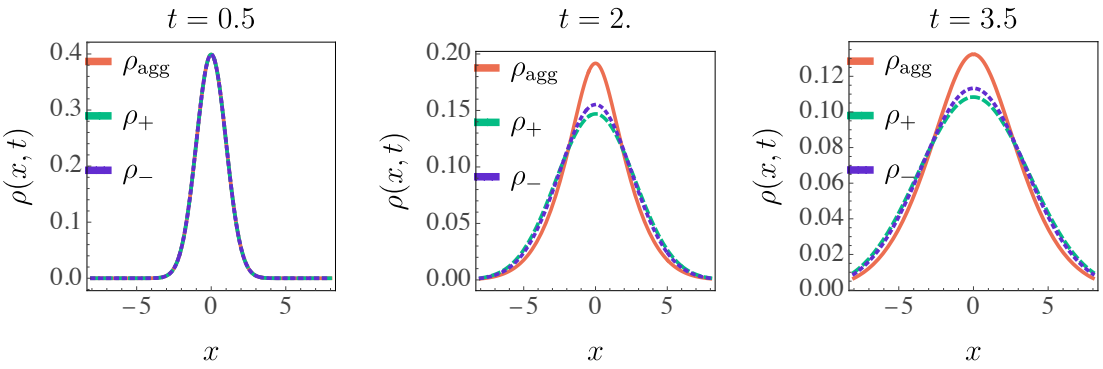


Figure 3.6: Diffusion-aggregation behaviour type I - Evolution of Gaussian distributions according to the chemotaxis models, Eqs. (3.23) and (3.33), at different times. In this example, we chose $\kappa = 3$ and $\omega = 2$ as for the model (3.33), these values would correspond to a scenario with high segregation.

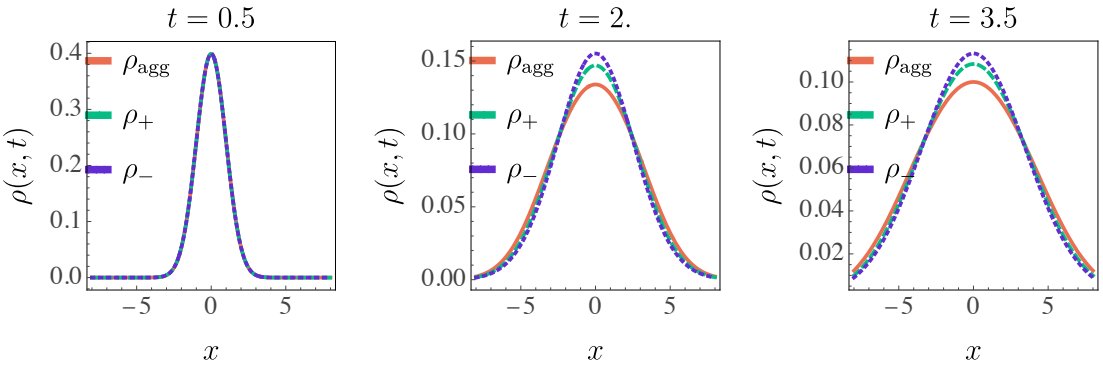


Figure 3.7: Diffusion-aggregation behaviour type II - Similar to the previous example, but fixing $\kappa = -3$. The diffusion described by the model (3.33) spreads faster than the other two cases.

3.4.3 Ultrafast Electron Transport

The scanning ultrafast electron microscopy is a technique efficiently implemented to observe transient super-diffusion behaviour in the dynamics of electrons and holes in Si after excitation with a short pulse laser[55]. This configuration allows observing an ultrafast electron transport phenomenon, which worths our attention in this section.

As we will see, those early electronic excitation stages are significantly influenced by the presence of nonlinearities that are beyond the scope of ordinary diffusion models. In our approach, we shall introduce modifications to one of our equations (3.30) and compare it with a phenomenological model, which formidably coincides with the experiment.

Let μ be the electron mobility in the transport of excited carriers with fundamental charge e. Then the Einstein relation[73] becomes $D_0 = \mu k_B T_0/e$, where k_B is the Boltzmann constant and T_0 is the room temperature (circa 300K). Whenever the excess energy of excited carriers is taken into account, a transient-diffusion term will emerge:

$$D_+(t) = \frac{\mu k_B T^*(0)}{e} \exp_+\left(-\frac{t}{\tau}\right) ,$$

where $T^*(0)$ is the initial carrier temperature ($T^*(0) \gg T_0$), τ is the relaxation time[38] and $\exp_+$ is a stretched exponential function, see Appendix B. In fact, one is aware that $\exp_+(-t/\tau)$ decays slower than $\exp(-t/\tau)$, although both vanish after a brief relaxation time τ.

Accordingly, we shall explore a transient-diffusion version of Eqs. (3.30) by means of the substitution $D \to D_0 + D_+(t)$. As a remark, in the present discussion, we shall refrain from considering the case related to D_- since it performs transient-diffusion slower than the phenomenological model of Ref.[55]:

$$\partial_t \rho_c = (D_0 + D^*(t))\nabla^2 \rho_c \, , \tag{3.34}$$

ρ_c denotes the distribution associated with the carriers and

$$D^* = \frac{\mu k_B T^*(0)}{e} \exp\left(-\frac{t}{\tau}\right) .$$

We solved numerically Eqs. (3.30) and (3.34) implementing the initial data used in Ref.[55], namely, a relaxation time $\tau = 77$ps and initial carrier temperature $T^*(0) = 4 \times 10^5$K for electrons, and $\tau = 161$ps and $T^*(0) = 2.7 \times 10^5$K for holes. Here we set

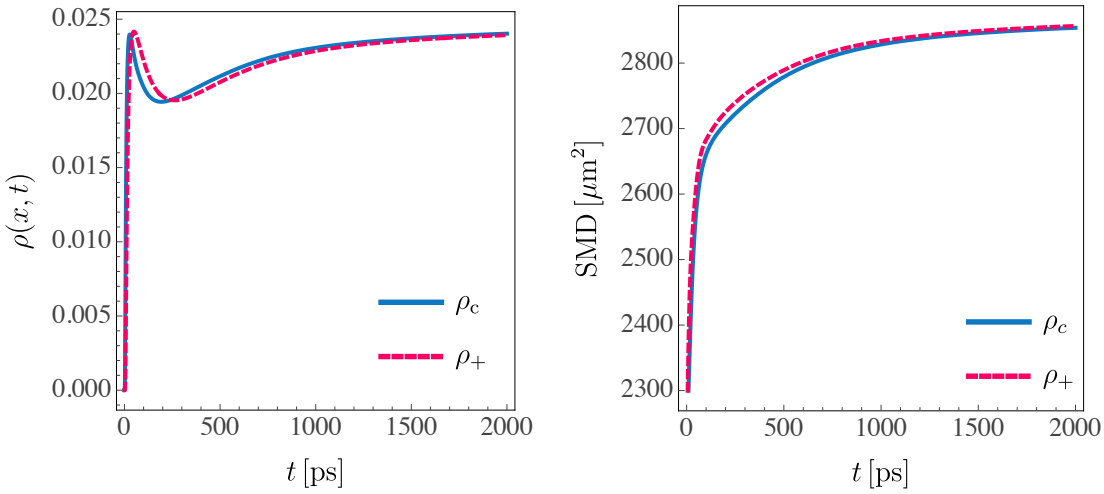


Figure 3.8: Transient-diffusion in electron transport - Left panel: It is shown the normalised distributions characterising the transport of excited carriers. Right panel: Square mean deviation of transient diffusion regarding non-interacting carriers ρ_c and interacting carriers ρ_+.

these same values for Eq. (3.34) except we chose different relaxations times for Eq. (3.30), since our model already incorporates those attributes related to nonequilibrium phenomena.

In the case of holes, we have considered an initial temperature $T^*(0) = 2.16 \times 10^5$K and a relaxation time $\tau = 40$ns, i.e. more than three hundred times the relaxation time adjusted phenomenologically for Eq. (3.34). The results are shown in Fig. 3.8. Notice the behaviour of both distributions—left panel—as well as the square mean deviation (SMD)—right panel—are roughly the same. The right panel exhibits a transient superdiffusion that increases monotonically in both models at early times, less than 250ps, whereas a steady-diffusion state is attained as time elapses.

The diffusion behaviour described by Eq. (3.30) is a consequence of its high relaxation time as compared with those values of τ given for the model in Ref.[55]. The reason is that the relaxation parameter τ associates with the time that the sample needs to reach the equilibrium with the medium after the laser pulse excitation.

Recall that Eq. (3.30) comes from a nonequilibrium background, involving nonlinear forces and other interactions, which have been disregarded in (3.34). Therefore, our model inherently needs a larger time to reach the equilibrium. Yet, if higher laser intensities feed the sample, the presence of electron-electron interactions would not be

negligible anymore and the description brought by Eq. (3.34) might fall into controversial results. Indeed, our model carries enough flexibility to describe a nonequilibrium stage at prior times, converging onto the region of steady-states for large periods of relaxation time.

4

Optimisation in Information Theory

The inclusion of non-extensive entropies in information theory has gained recognition in recent decades. Nonetheless, those parameter-dependent entropies frequently receive more attention than those depending exclusively on the probability distribution. In this chapter we study a set of measures of lengths for prefix-free codes based on the non-parametric, non-extensive entropies S_+ and S_-. We will show that, by implementing a variational method, one is able to find the optimal codeword lengths that permit formulating generalised noiseless coding theorems associated with each of these measures of entropy, implying the resulting average lengths are bounded from below and above.

If our scheme truncates to first-order, Shannon's framework is promptly recovered. Otherwise, there exists a regime where our proposal might produce more efficient data compression than that the standard theory estimates. Besides, we shall discuss how to apply the variational method to attain optimal lengths given generic measures of entropy.

4.1 The Standard Results

In 1948, Claude Shannon published the first rigorous study on data compression[66], setting the modern foundations of information theory. He analysed an alphabet of D symbols produced by a collection of uncorrelated random sources of letters to quantify to what extent could one compress an arbitrary number of emitted symbols. Shannon

found the entropy itself limits any process of data compression, such that a codeword is optimal if its average length attains the entropy. In other words, this result led to quantify the average amount of information—or absence of redundancy—by simply measuring the entropy.

However, as we have seen in previous chapters, there are numerous measures of entropy, in such a way that one would think data compression could reach absurdly small sizes according to the entropy of our interest. In principle, it is possible as long as the measure of entropy is stable and fulfils at least SK1-SK3. As far as we know, there are only three measures of entropy that satisfy such conditions: S_+, S_- and S_{BG}. Even the Rényi entropy S_α fails to be stable, despite it belongs to the equivalence class $(c, d) = (1, 1)$, hence any compression process leveraged by S_α must be cautiously examined.

In what follows, we shall slightly modify the notation that we have formerly adopted to refer to measures of entropy. For the sake of consistency with the notation usually found in the literature, henceforth we shall represent the entropy with the letter H instead of S, used in physics. Moreover, in this context the Boltzmann-Gibbs entropy is known as the Shannon's entropy:

$$H_D^S(P) = -\sum_j^\Omega p(x_j) \ln_D(p(x_j)) \ ,$$

where D is the dimension of the alphabet and Ω the number of states. In information theory, one used to draw on an alphabet of two symbols, 1s and 0s—known as bits—hence $D = 2$. Yet we shall stick to the general case throughout our discussion unless another thing is properly specified.

Likewise, the nonextensive entropies S_+ and S_- are now expressed as:

$$H_D^+(P) = -\sum_j^\Omega p(x_j) \ln_D^+(p(x_j)) \ , \quad H_D^-(P) = -\sum_j^\Omega p(x_j) \ln_D^-(p(x_j)) \ , \tag{4.1}$$

where $\ln^\pm$ are generalised logarithms (see Appendix B).

The entropies (4.1) demand another generalised quantities, as we will see below. For instance, we cannot find effective coding theorems in terms of these measures if the constraint

$$\sum_j^\Omega D^{-l(x_j)} \leq 1 \ ,$$

is imposed to minimise the corresponding codeword length, therefore, leading to the necessity of a generalised constraint, as stated in Prop. 1.

Intuitively, one could ask to what degree could the standard information theory[26] be modified by implementing the information measures (4.1), given Fig. 2.1 tells in advance the existence of upper and lower bounds on the Shannon's entropy in certain regions of Ω.

In turn, we shall present coding theorems in terms of the non-extensive entropies (4.1). We will apply a variational method to find the average length L that a codeword is allowed to attain when the data compression is subject to a given constraint. This constraint is not arbitrary. As we will show, there is an intertwining between the measure of entropy and the minimisation constraint so that the entropy preserves its functional structure.

For the sake of simplicity, we will assume the communication channel is perfectly noiseless, that is the information between two points A and B is transmitted without any loss. In addition, we will suppose those transmitted codes are prefix-free, i.e. there is no a single codeword which is the initial segment of any other available codeword in the set generated by a random source X.

The purpose of our discussion is to find codes that minimise a monotonic increasing function of the form

$$L = \varphi^{-1}\left(\sum_{j}^{\Omega} p(x_j)\varphi(l(x_j))\right), \tag{4.2}$$

where $p(x_j)$ is the probability of finding the codeword x_j, $l(x_j) \in \mathbb{Z}_+$ is a function that relates a codeword x_j to its length, and $\varphi(l(x_j)) : \mathbb{R}_+ \to \mathbb{R}_+$ is the Nagumo-Kolmogorov function, which specifies the cost of managing a sequence of length $l(x_j)$.

The quantity defined in (4.2) is called the *average length for the cost function* φ, in accordance with Campbell[12], although for brevity we refer to it simply as the average or mean length.

A number of cost functions φ have been studied in Ref.[2]. Still for most generic entropies, we simply have that

$$\varphi(z) = z\ .$$

In the case of Shannon's entropy, for instance, the average length

$$\begin{aligned} L_S &= \varphi^{-1}\left(\sum_j^{\Omega} p_S(x_j)\varphi(l_S(x_j))\right) \\ &= \sum_j^{\Omega} p_S(x_j) l_S(x_j) \end{aligned} \tag{4.3}$$

is optimised via the functional

$$J_S = L_S + \gamma K_S \ , \tag{4.4}$$

where $\gamma \in \mathbb{R}$ is a Lagrange multiplier and the constraint

$$K_S = \sum_j^{\Omega} D^{-l_S(x_j)} \leq 1 \ , \tag{4.5}$$

is the Kraft-McMillan inequality[22], accounting for the fundamental control function in the optimisation of this case.

To solve the optimisation problem in (4.4), one has to vary J_S with respect to $l_S(x_j)$ and equate to zero, recalling that the rule to find the extrema of (4.4) must be the same for every $x_j \in X$. We readily outperform the calculation to obtain:

$$l_S^*(x_j) = -\ln_D(p_S(x_j)) \ , \tag{4.6}$$

the notation * indicates that the average length is optimal, to wit, it solves the problem (4.4).

By substituting l_S into (4.3) we recover Shannon's entropy,

$$\begin{aligned} L_S &= \sum_j^{\Omega} p_S(x_j) l_S(x_j) \\ &= -\sum_j^{\Omega} p_S(x_j) \ln_D(p_S(x_j)) \\ &= H_D^S(P) \ , \end{aligned} \tag{4.7}$$

which means the measure of entropy determines the average optimal length. This is the original result derived by Shannon[66] which places entropy as a fundamental measure of information, in that it quantifies the length a codeword should have to achieve its optimal compression without loss of information.

4.2 Effective Noiseless Coding Theorems

In turn, we are to proceed in an analogous way but using the measures of entropy (4.1). Nonetheless, a generalisation to (4.5) shall be devised to optimise the mean lengths related to $H_D^{\pm}(P)$.

This is a striking aspect of our scheme, for the quantity to be minimised will grant a different rate of data compression than that estimated in the standard theory. Hence, in our case, the problem consists of choosing codes that become minima when they are subject to the quantities in the following definition.

Proposition 1. *The entropies $H_D^{\pm}(P)$ define functionals of the class (4.4)*

$$J_{\pm} = L_{\pm} + \gamma K_{\pm} \ , \tag{4.8}$$

where γ is a Lagrange multiplier and the average lengths $L_{\pm}$ are univocally determined by the constraints:

$$K_{\pm} = \sum_{j}^{\Omega} \sum_{j}^{\infty} a_{\pm}(j) \Gamma \left[j+1, -\ln D^{-l_{\pm}(x_j)} \right] \leq const. \ , \tag{4.9}$$

the individual lengths $l_{\pm}(x_j)$ will eventually be related to the probability distributions $p_{\pm}(x_j)$, the real coefficients $a_{\pm}(j)$ are given in Appendix B and

$$\Gamma(y,x) = \int_x^{\infty} dz \, z^{y-1} e^{-z} \ ,$$

is the incomplete gamma function.

Proof. Since we are able to express (4.1) in the generalised form, the corresponding Nagumo-Kolmogorov function is $\varphi(z) = z$. Consistently, the average lengths in (4.8) read as

$$L_{\pm} = \sum_{j}^{\Omega} p_{\pm}(x_j) l_{\pm}(x_j) \ . \tag{4.10}$$

To find the optimal individual lengths $l_{\pm}^{*}(x_j)$, we now differentiate (4.8) with respect to $l_{\pm}(x_j)$, that is:

$$\frac{\partial J_{\pm}}{\partial l_{\pm}(x_j)} = \frac{\partial L_{\pm}}{\partial l_{\pm}(x_j)} + \gamma \frac{\partial K_{\pm}}{\partial l_{\pm}(x_j)} \ , \tag{4.11}$$

provided we are looking for a global minimum, the equality must vanish term to term, therefore

$$\begin{aligned}\frac{\partial L_{\pm}}{\partial l_{\pm}(x_j)} &= \sum_{j}^{\Omega} p_{\pm}(x_j) \\ &= -\gamma \frac{\partial K_{\pm}}{\partial l_{\pm}(x_j)} \\ &= \gamma \ln D \sum_{j}^{\Omega} \exp^{\pm}\left[-l_{\pm}(x_j)\ln D\right] \\ &= \sum_{j}^{\Omega} \exp^{\pm}\left[-l_{\pm}(x_j)\ln D\right] \ ,\end{aligned} \tag{4.12}$$

where the Lagrange multiplier has been selected as $\gamma = \frac{1}{\ln D}$ and $\exp^{\pm}$ are stretched exponential functions (see Appendix B).

In addition, the equality in (4.12) is satisfied for all $x_j \in X$ iff

$$l_{\pm}^{*}(x_j) = -\ln_D^{\pm}\left(p_{\pm}(x_j)\right) \ .$$

At this point it should be evident that we have obtained two optimal individual lengths as given by each probability distribution $p_{\pm}(x_j)$, for that reason we have appended the label $\pm$ to $l_{\pm}^{*}(x_j)$, since the lengths and the distributions are univocally related. That completes the proof. □

Remark 6. *The non-extensive entropies $H_D^{\pm}$ bound the expected lengths $L_{\pm}$ from above and below, which is consistent with the standard formulation. Thus, we are entitled to introduce the coding Theorems 1 and 2. However, before formally introducing these theorems, some comments deserve attention:*

1. *Note that the individual lengths $l_{\pm}$ and the probabilities $p_{\pm}$ are determined by each other, therefore the constraints (4.9) are not arbitrary, but must be constructed so that measures of entropy are recovered at the optimal points.*

2. *We also remark the standard theory is straightforwardly recoverable from our scheme. Notice the Kraft inequality (4.5) is obtained by truncating $K_{\pm}$ at first*

order, that is:

$$\begin{aligned} K_{\pm} = a_{\pm}(0) \sum_{j}^{\Omega} D^{-l_{\pm}(x_j)} \\ + a_{\pm}(1) \sum_{j}^{\Omega} \Gamma[2, -\ln[D^{-l_{\pm}(x_j)}] \\ + a_{\pm}(2) \sum_{j}^{\Omega} \Gamma[3, -\ln[D^{-l_{\pm}(x_j)}] + \cdots , \end{aligned} \tag{4.13}$$

with $a_{\pm}(0) = 1$.

Equivalently, the limit $\Omega \to \infty$ means the system will possess low probabilities due to a high number of accessible states, in that case our proposal is asymptotically equivalent to Shannon's theory, symbolically expressed as $p_S(x) = \lim_{\Omega\to\infty} p_{\pm}(x)$.

Theorem 1. *Let D be the number of symbols in an alphabet, then the expected lengths defined by Eq. (4.10) and the entropies $H_D^{\pm}(P)$ satisfy the relation*

$$L_{\pm} \geq H_D^{\pm}(P) \, ,$$

with equality iff $l_{\pm}^*(x_j) = -\ln_D^{\pm}(p_{\pm}(x_j))$ for every x_j in X.

Proof. We begin by writing the difference between the expected lengths $L_{\pm}$ and the entropies $H_D^{\pm}$, we readily obtain:

$$\begin{aligned} L_{\pm} - H_D^{\pm}(P) &= \sum_{j}^{\Omega} p_{\pm}(x_j) l_{\pm}(x_j) \\ &\qquad + \sum_{j}^{\Omega} p_{\pm}(x_j) \ln_D^{\pm}(p_{\pm}(x_j)) \\ &= \sum_{j}^{\Omega} p_{\pm}(x_j) \left[l_{\pm}(x_j) + \ln_D^{\pm}(p_{\pm}(x_j)) \right] \\ &\geq 0 \, , \end{aligned} \tag{4.14}$$

necessarily leading to

$$l_{\pm}(x_j) \geq -\ln_D^{\pm}(p_{\pm}(x_j)) \ ,$$

given that every $l_{\pm}(x_j)$ is an integer. Then, the equality is attained iff the individual lengths $l_{\pm}(x_j) = l_{\pm}^*(x_j)$ are optimal. And the theorem is demonstrated. □

Furthermore, the entropies $H_D^{\pm}$ amount to a lower bound on the expected lengths $L_{\pm}$, yet as we are to show these lengths are within one *dit* of the lower bound as well.

Theorem 2. *Consider an alphabet of D symbols and a random source distribution X. Let $l_{\pm}(x_j)$ be the optimal individual lengths that solve the optimisation problem (4.8), where the associated average lengths are defined by Eq. (4.10). Then, the following relation is true:*

$$H_D^{\pm}(P) \leq L_{\pm} < H_D^{\pm}(P) + 1 \; . \tag{4.15}$$

Proof. According to Theorem 1, the simple choice of codeword lengths

$$l_{\pm}^{*}(x_j) = -\ln_D^{\pm}\left(p_{\pm}(x_j)\right)$$

leads to $L_{\pm}^{*} = H_D^{\pm}$. Notwithstanding, to assure that every $l_{\pm}(x_j)$ is an integer, we shall take

$$l_{\pm}(x_j) = \left\lceil -\ln_D^{\pm}\left(p_{\pm}(x_j)\right)\right\rceil$$

in order that the individual lengths fulfil

$$-\ln_D^{\pm}\left(p_{\pm}(x_j)\right) \leq l_{\pm}(x_j) < -\ln_D^{\pm}\left(p_{\pm}(x_j)\right) + 1 \; . \tag{4.16}$$

Thereupon, we multiply each member by $p_{\pm}(x_j)$ and sum over all x_j, which leads to

$$\begin{aligned} -\sum_j^{\Omega} p_{\pm}(x_j)\ln_D^{\pm}\left(p_{\pm}(x_j)\right) &\leq \sum_j^{\Omega} p_{\pm}(x_j) l_{\pm}(x_j) \\ &< -\sum_j^{\Omega} p_{\pm}(x_j)\ln_D^{\pm}\left(p_{\pm}(x_j)\right) \\ &\quad + \sum_j^{\Omega} p_{\pm}(x_j) \; , \end{aligned} \tag{4.17}$$

then from Eqs. (4.1) and (4.10) together with the normalisation of the probability, we finally arrive at the expression

$$H_D^{\pm}(P) \leq L_{\pm} < H_D^{\pm}(P) + 1 \; , \tag{4.18}$$

and we have the theorem. □

We would like to draw attention that the expected lengths $L_{\pm}$ satisfy

$$H_D^{\pm} \leq L_{\pm} < H_D^{\pm} + 1 \; ;$$

nonetheless, those optimal codes prescribed by $H_D^{\pm}$ can only be better than those prescribed by $L_{\pm}$, therefore one addresses Theorem 2.

To clarify how our formulation works, we generated some random processes with uncorrelated sources that we shall use as datasets to compute the average lengths $L\pm$ as well as L_S, cf. Figure 4.1 and Table 4.1.

Events	20	150	300	600
L_S^* [bits]	4.60	7.42	8.45	9.45
L_+^* [bits]	4.10	7.30	8.38	9.40
L_-^* [bits]	4.75	7.49	8.51	9.47

Table 4.1: Chart of optimal lengths varying the number of random events, or number of codewords composing a dataset.

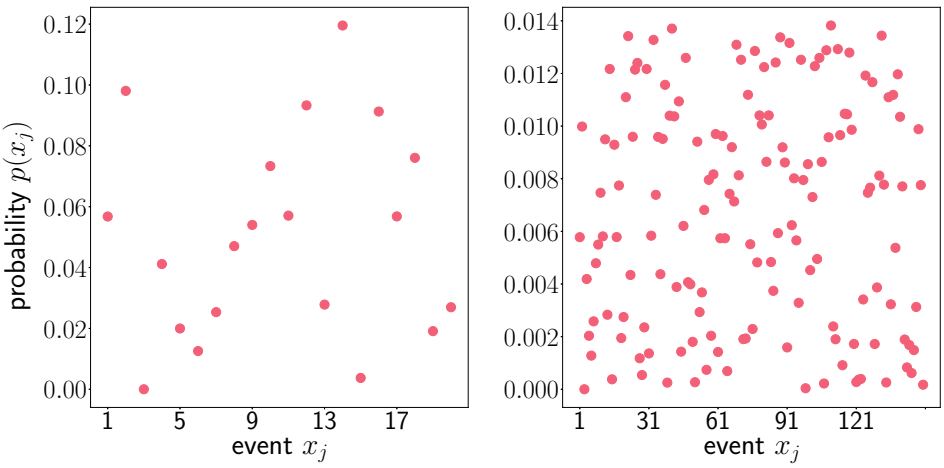


Figure 4.1: Uncorrelated random processes - Left panel: A random process with 20 events. Right panel: A random process with 150 events. The x axis represents the individual events x_j whose probabilities $p(x_j)$ were randomly assigned.

Considering the random process in Fig. 4.1, with a binary alphabet $D = 2$, we have obtained an average length L_S^*=4.6bits, whereas the average lengths regarding

entropies H^+ and H^- are L^*_+=4.1bits and L^*_-=4.75bits, respectively. That means that a more efficient transmission process would result from the entropy H^+ in comparison to a code compressed via H^S. However as the number of random events increases, the lengths $L^*_\pm$ tend to coincide with L^*_S. For instance, with respect to the process with 150 events in Fig. 4.1, the standard average length is L^*_S=7.42bits, while L^*_+=7.3bits and L^*_-=7.49bits, hence diminishing the difference between L^*_S and $L^*_\pm$, see also Table 4.1: The more random events the less difference between the three average lengths. This comes as no surprise since entropies $H^\pm$ converge asymptotically to H^S.

4.3 Effective Channel Capacities

It would appear that to increase the capacity of a communication system, it would be sufficient to increase the number of signalling events that move freely from Alice to Bob. That is, why not arbitrarily transmit various packets of voltages per symbol over a communication channel? In practice, this conveys unfavourable consequences. As Alice increases the number of emitted signals, Bob is left with other difficulties. The reason is all the signals Alice casts are constrained by the unavoidable surrounding noise. So that Bob retrieves the proper signals, the difference between the signal events sent by Alice must be higher than the noise level.

To skirt this problem, the notion of *channel capacity* becomes fundamental to know the maximum amount of symbols per unit time, n, that is emitted and the many differences per symbols, s, that can be selected for the message space s^n. Indeed, those codewords Bob computes generate a probability distribution living in the message space that the decision tree s^n spans.

Definition 8. *Let P and Q be different probability distributions. The amount of information shared between them is measured through the mutual information*

$$I(P,Q) = H_D(Q) - H_D(Q|P)\ ,$$

where $H_D(Q|P)$ is the conditional entropy. Then the channel capacity is defined as:

$$C = \max_P I(P,Q). \tag{4.19}$$

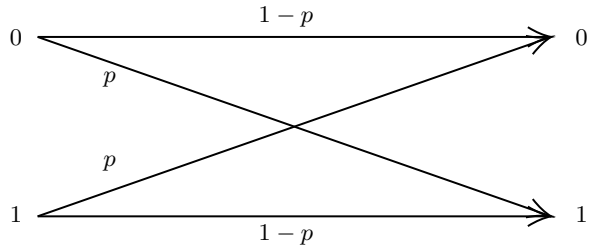
This is the standard definition of channel capacity and is interpreted as the maximum rate at which the signal is reconstructed in terms of a given measure of entropy.

We shall proceed considering the entropies (4.1), which bring novel bounds on the channel capacity due to Shannon.

4.3.1 Binary Symmetric Channel

In some circumstances we obtain the same outcome given two different inputs, making the signals looking ambiguous. As a first approach to trim this problem, a possible choice of unambiguous inputs is realisable with a binary symmetric channel (BSC), which allows the signal reconstruction with a negligible rate of error.

The BSC is one the simplest models of communication channels with errors. Each input is complemented with probability p, see the diagram below.



If an error occurs, an input with value 1 (0) will be regarded as 0 (1), undermining our ability to identify those error bits and leaving us with set of non-trusted messages. For that reason, we shall assume that every bit Alice sends has a negligible probability of error. Henceforth all operations are binary, i.e. $D = 2$, so we prefer to drop this label from our notation.

To calculate the channel capacity of a BSC, first note the mutual information $I(P, Q)$ is bounded by[22]

$$I(P, Q) \leq 1 - H(P) \,, \tag{4.20}$$

observe that the equality is attained whenever a uniform distribution feeds the input. Accordingly, from (4.19) we easily obtain

$$C_{\text{BSC}} = 1 - H(P) \,, \tag{4.21}$$

and we have the general expression for the channel capacity of a BSC.

In particular, note that if Shannon's entropy quantifies the channel capacity (4.21) of a BSC communication system, where $P = \{p, 1-p\}$, it yields

$$C^{S}_{\text{BSC}} = 1 + p \ln p + (1-p) \ln(1-p) \,, \tag{4.22}$$

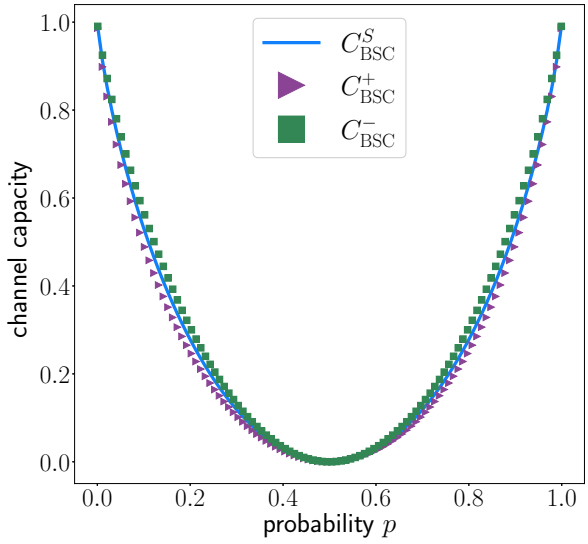


Figure 4.2: Capacity of a BSC - The figure shows the capacity of a BSC in terms of the measures of entropy $H^+(P), H^-(P)$ and $H^S(P)$. The entropy $H^-(P)$ estimates a hardly greater capacity than that obtained from Shannon's entropy.

where logarithms are base 2.

We proceed in a similar fashion to determine the channel capacities of a BSC in terms of the entropies $H^{\pm}(P)$, namely:

$$C^+_{\mathrm{BSC}} = \frac{\sqrt{2} - p^p - (1-p)^{1-p}}{\sqrt{2} - 2} \,, \tag{4.23}$$

and

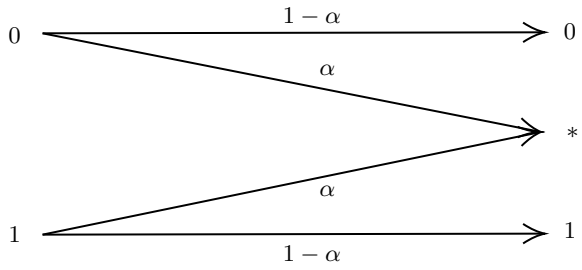
$$C^-_{\mathrm{BSC}} = \frac{2\sqrt{2} - p^{-p} - (1-p)^{-(1-p)}}{2\sqrt{2} - 2} \,, \tag{4.24}$$

where we normalised the capacities $C^{\pm}_{\mathrm{BSC}}$ to be comparable with the standard estimation C^S_{BSC}.

The capacities $C^+_{\mathrm{BSC}}, C^-_{\mathrm{BSC}}$ and C^S_{BSC} have been plotted in Fig. 4.2. Observe that $C^-_{\mathrm{BSC}} \geq C^S_{\mathrm{BSC}}$ with equality at the points $p = \{0, \frac{1}{2}, 1\}$, otherwise C^-_{BSC} exhibits a modest improvement with respect to C^S_{BSC}, indicating a tentative upper limit above Shannon's. Consistently, at $p = \frac{1}{2}$ the three capacities coincide, since at this point occurs the highest degree of uncertainty, resembling the scenario in which Bob cannot form any judgment from the received bits.

4.3.2 Binary Erasure Channel

We shall survey right away another scenario in which a fraction α of bits is erased or lost during the transmission process, yet the receiver is aware of those bits. This is known as a binary erasure channel (BEC). As for the BSC, there are two entries, but now there will be three exits:



This diagram represents an input X, which emits 0s and 1s with probabilities p and $1-p$, correspondingly. Whereas the output delivers 0s, 1s or $*$s with probabilities $(1-\alpha)p$, $(1-\alpha)(1-p)$ or α, in that order. Furthermore, if the source emits either 0s or 1s, the probabilities of receiving these bits without errors are $(1-\alpha)p$ and $(1-\alpha)(1-p)$, otherwise the probabilities that these bits swap across the communication channel are $p\alpha$ and $(1-p)\alpha$.

The channel capacity of a BEC with an erasure probability α, is calculated as:

$$C_{\text{BEC}} = \max_{P} H(X) + H(Y) - H(X,Y), \tag{4.25}$$

where $H(X,Y)$ is the joint entropy computed in terms of the joint distributions, which alternatively can be done via transition matrices[39].

Once more, we start by computing the capacity (4.25) in terms of Shannon's entropy, we obtain:

$$C^{S}_{\text{BEC}} = \max_{P} (1-\alpha) H^{S}(P), \quad P = \{p, 1-p\}, \tag{4.26}$$

since the entropy attains its maximum for a uniform distribution, $p = \frac{1}{2}$, the channel capacity of the BEC reduces to $C^{S} = 1-\alpha$, a well known result in Shannon's theory[26].

It shall not escape attention, that the channel capacities (4.22) and (4.26) are completely additive, that is, given the distributions p_1 and p_2, we have:

$$C^{S}(p_1 p_2) = C^{S}(p_1) + C^{S}(p_2)\ , \tag{4.27}$$

a direct result that follows from

$$C^S(p_1 p_2) = \max_{p_{X_1,X_2}} I(X_1, X_2; Y_1, Y_2) \ , \tag{4.28}$$

with

$$I(X_1, X_2; Y_1, Y_2) = I(X_1, Y_1) + I(X_2, Y_2) \ . \tag{4.29}$$

Although this property will not be satisfied by the non-extensive entropies $H^{\pm}(P)$, in which case the additivity is asymptotically achieved, rather we are not interested in such aspects but precisely in the non-additive consequences and possible applications, if any.

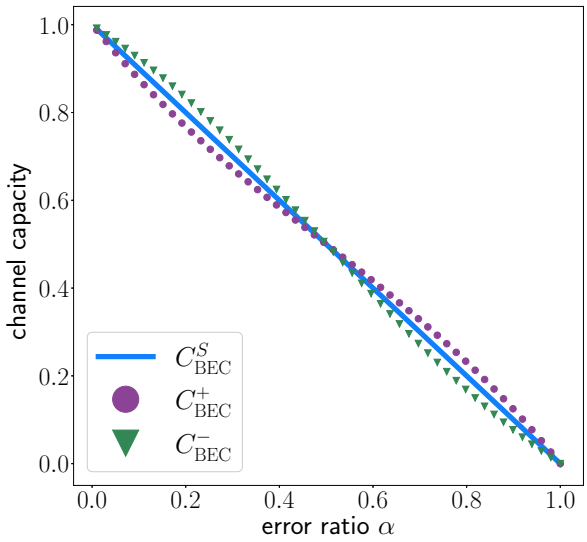


Figure 4.3: Capacity of a BEC - Channel capacity of a BEC, for measures of entropy that depend only on the distribution P, namely: $H^+(P), H^-(P)$ and $H^S(P)$.

We shall now compute the channel capacity of a BEC addressing the information measures (4.1). Similar to the BSC case, we will have to normalise the corresponding capacities to compare with the standard result (4.25). We get:

$$C^+_{\mathrm{BEC}} = \max_P \frac{1 - \alpha^\alpha + H^+(P) - H^+(P\alpha)}{2 - \sqrt{2}}, \tag{4.30}$$

and

$$C^-_{\mathrm{BEC}} = \max_P \frac{\alpha^{-\alpha} - 1 + H^-(P) - H^-(P\alpha)}{2\sqrt{2} - 2}, \tag{4.31}$$

as it happens for Eq. (4.26), here the maximum will be attained at $p = \frac{1}{2}$.

The channel capacities C^{S}_{BEC} and $C^{\pm}_{\text{BEC}}$ have been plotted in Fig. 4.3. Interestingly $C^{\pm}_{\text{BEC}}$ behave such that the three measures coincide at the points $p = \{0, \frac{1}{2}, 1\}$, although the character of $C^{\pm}_{\text{BEC}}$ flips at $p = \frac{1}{2}$, which shows the flexibility of our approach.

We interpret the flipping behaviour of $C^{\pm}_{\text{BEC}}$ as follows. Note that C^{-}_{BEC} will establish an upper bound on C^{S}_{BEC} as long as the fraction of errors α does not dominate the communication channel, i.e. $< 50\%$. Otherwise there is a tradeoff and C^{-}_{BEC} will be slightly reduced immediately after the threshold at $p = \frac{1}{2}$. For the same reason, the lower bound suggested by C^{+}_{BEC}, will eventually be promoted to an upper bound, with respect to C^{S}_{BEC}, when the ratio α dominates the channel ($> 50\%$).

4.4 Parametric Average Lengths

The application of the variational method discussed in Secs. 4.1 and 4.2 is not limited to those entropies of the form (2.2). The same method works in those entropies such as the Rényi entropy. Below we are to illustrate how this method is applied to the Tsallis and Rényi entropies to find their optimal lengths. However, we will not derive their corresponding coding theorems, which have been studied in depth in Refs.[11,60], although following an approach different than that exposed here.

4.4.1 Tsallis Average Length

The measure of entropy proposed in Refs.[37,75], known as the Tsallis entropy, is a parameter-dependent entropy of the form (2.2), it reads

$$H^{T}_{D}(P) = -\sum_{j}^{\Omega} p(x_j) \ln^{q}_{D}(p(x_j)), \quad q > 0, \tag{4.32}$$

where the q-logarithm function is defined as

$$\ln^{q}(x) = \frac{x^{q-1} - 1}{q - 1} \quad \text{if} \quad q \neq 1 \quad \text{and} \quad x > 0\,;$$

if $q = 1$ the standard logarithm function is retrieved and if $x \leq 0$ the function becomes undefined.

We are to show how the variational method is applied to find the optimal length L^{*}_{T} in the context of Tsallis statistics. Nonetheless, as it was discussed in Sec. 4.2, we need the proper constraint.

Proposition 2. *Let D be the number of symbols in an alphabet. For the measure of entropy $H_D^T(P)$ there exists a functional of the form:*

$$J_T = L_T + \gamma K_T \ , \tag{4.33}$$

where γ is a multiplier to determine and L_T is an average length univocally related to the entropy iff the constraint is defined as:

$$K_T = \sum_j^{\Omega} \left[1 + (1 - D^{q-1}) l_T(x_j)\right]^{\frac{1}{q-1}} \leq const. \ , \quad q > 0 \ , \tag{4.34}$$

where $l_T(x_j)$ is the length of the codeword x_j.

Proof. We shall prove that if L_T is optimised subject to the constraint (4.34), we are led to the Tsallis entropy. Since the entropy (4.32) is of the form (2.2), the Nagumo-Kolmogorov function is $\phi(z) = z$. Hence, we write down the functional (4.33) as:

$$J_T = \sum_j^{\Omega} p_T(x_j) l_T(x_j) + \gamma \sum_j^{\Omega} \left[1 + (1 - D^{q-1}) l_T(x_j)\right]^{\frac{1}{q-1}} \ , \tag{4.35}$$

varying J_T w.r.t. $l_T(x_j)$ and equating to zero, we have that for every x_j

$$\begin{aligned} p_T^q(x_j) &= \gamma \ln^q D \left[1 + (1 - D^{q-1}) l_T(x_j)\right]^{\frac{q}{q-1}} \\ &= \left[1 + (1 - D^{q-1}) l_T(x_j)\right]^{\frac{q}{q-1}} \ , \end{aligned} \tag{4.36}$$

having selected $\gamma = 1/\ln^q D$. Elevating both members to $q - 1$:

$$p_T^{q-1}(x_j) = 1 + (1 - D^{q-1}) l_T(x_j) \ . \tag{4.37}$$

We want to solve for $l_T(x_j)$ from the last equation. To proceed, notice this individual length must be optimal at this point, therefore we get:

$$l_T^*(x_j) = -\frac{p_T^{q-1}(x_j) - 1}{D^{q-1} - 1} = -\ln_D^q(p_T(x_j)) \ , \tag{4.38}$$

finally, substituting $l_T^*(x_j)$ into the average length:

$$\begin{aligned} L_T^* &= \sum_j^{\Omega} p_T(x_j) l_T^*(x_j) \\ &= -\sum_j^{\Omega} p_T(x_j) \ln_D^q(p_T(x_j)) \\ &= H_D^T(X) \ , \end{aligned} \tag{4.39}$$

thus, recovering the entropy (4.32). That concludes the proof. □

Remark 7. *Looking at the constraint* (4.34)*, if* $q \to 1$*, then* $K_T \to K_S = \sum_j^{\Omega} D^{-l_S(x_j)}$*, and Eq.* (4.5) *is retrieved.*

The variational method is applied to obtain the optimal expected lengths in view of a given measure of entropy and subject to an adequate constraint. This constraint is not arbitrary since the optimal length must correspond to the entropy, on account of the optimal individual lengths $l^*(x_j)$. For a code to be minimised in view of the Tsallis statistics, the suitable constraint to impose is (4.34), so that coding theorems can be stated.

Following an algebraic method, the authors in Ref.[60] proved the bounds on L_T are given in terms of the measure of entropy (4.32), leading to the corresponding coding theorems in Tsallis statistics. We recall that $H_D^T(P)$ is unstable.

4.4.2 Rényi Average Length

We shall discuss the corresponding average lengths to the Rényi entropy[64]

$$H_D^R(X) = \frac{1}{1-\alpha} \ln_D \left(\sum_j^{\Omega} p^{\alpha}(x_j) \right), \quad \alpha > 0 \,. \tag{4.40}$$

As we discussed in Chapter 2, this entropy does not adopt the form (2.2), rather it belongs to the family of entropies expressed as

$$H_D = \phi_D \left(\sum_j^{\Omega} g(p_j) \right) ,$$

see Remark 1.

Proposition 3. *For the measure of entropy* $H_D^R(P)$ *there exists a functional of the form:*

$$J_R = L_R + \gamma K_R \,, \tag{4.41}$$

where γ *is a multiplier to determine and* L_R *is an average length univocally related to the entropy iff the constraint is defined as:*

$$K_R = \sum_j^{\Omega} D^{-l_R(x_j)} \,, \tag{4.42}$$

where $l_R(x_j)$ *is the length of the codeword* x_j*.*

Proof. To apply the variational method, we must choose the Nagumo-Kolmogorov function $\varphi(l(x_j)) = D^{\beta l_R(x_j)}$, with $\beta > 0$. Therefore, from Eq. (4.2), the length to optimise reads[11]

$$L_R = \frac{1}{\beta} \ln_D \left(\sum_j^{\Omega} p_R(x_j) D^{\beta l_R(x_j)} \right) , \tag{4.43}$$

substituting into Eq. (4.41), we obtain:

$$J_R = \frac{1}{\beta} \ln_D \left(\sum_j^{\Omega} p_R(x_j) D^{\beta l_R(x_j)} \right) + \gamma \sum_j^{\Omega} D^{-l_R(x_j)} , \tag{4.44}$$

which we shall variate w.r.t. $l_R(x_j)$. Even more, provided we are looking for a global minimum, $\partial J_R / \partial l_R(x_j) = 0$ implies every term on the right-hand side of (4.44) must vanish independently, hence

$$\frac{p_R(x_j) D^{\beta l_R(x_j)}}{\psi_\beta} = \gamma \ln(D) D^{-l_R(x_j)}, \tag{4.45}$$

where

$$\psi_\beta = \sum_j^{\Omega} p_R(x_j) D^{\beta l_R(x_j)} \tag{4.46}$$

with $\gamma = 1/\ln D$.

To find the definite expression for the optimal lengths $l_R^*(x_j)$ comprised in the relation (4.45), we consider Hölder's inequality:

$$\left(\sum_j^{\Omega} p(x_j) D^{\beta l(x_j)} \right)^{-\frac{1}{\beta}} \left(\sum_j^{\Omega} p^\alpha(x_j) \right)^{\frac{1}{1-\alpha}} \leq \sum_j^{\Omega} D^{-l(x_j)} \leq 1 , \tag{4.47}$$

with the condition $\alpha = (1+\beta)^{-1}$. Notice the sum in the first factor on the left-hand side is exactly ψ_β. Accordingly, we get:

$$\psi_\beta \leq \left(\sum_j^{\Omega} p_R^\alpha(x_j) \right)^{\frac{1}{\alpha}} , \tag{4.48}$$

and from Eq. (4.45) we also find

$$\psi_\beta = p_R(x_j) D^{(1+\beta) l_R(x_j)} , \tag{4.49}$$

replacing this expression into (4.48), yields

$$p_R(x_j) D^{(1+\beta) l_R(x_j)} \leq \left(\sum_j^{\Omega} p_R^\alpha(x_j) \right)^{\frac{1}{\alpha}} , \tag{4.50}$$

taking the logarithm on both sides and dividing by $\ln D$, one obtains

$$l_R(x_j) + \alpha \ln_D p_R(x_j) \leq \ln_D \sum_j^{\Omega} p_R^{\alpha}(x_j) \,, \tag{4.51}$$

with equality iff $l_R(x_j)$ is the optimal length $l_R^*(x_j)$, therefore, we obtain the optimal codeword lengths:

$$l_R(x_j)^* = -\ln_D \left(\frac{p_R^{\alpha}(x_j)}{\sum_j^{\Omega} p_R(x_j)^{\alpha}} \right), \tag{4.52}$$

coinciding with the result given in Ref.[11].

The next step is verifying that $l_R^*(x_j)$ is in truth optimal. One merely substitutes (4.52) into the mean length (4.43) to recover the Rényi entropy, that is:

$$\begin{aligned} \frac{1}{\beta} \ln_D \left(\sum_j^{\Omega} p_R(x_j) D^{\beta l_R(x_j)^*} \right) &= \frac{1}{\beta} \ln_D \left(\sum_j^{\Omega} p_R^{\alpha}(x_j) \right) \\ &\quad + \frac{1}{\beta} \ln_D \left(\sum_j^{\Omega} p_R^{1-\alpha\beta}(x_j) \right) \\ &= (1-\alpha) H_D^R(X) + \alpha H_D^R(X) \\ &= H_D^R(X) \,, \end{aligned} \tag{4.53}$$

and we conclude the proof. □

Remark 8. *Alternatively, since the optimal mean length L_R^* must coincide with the measure of entropy $H_D^R(X)$, it follows that Eq. (4.43) is also obtained through a map between the Tsallis and Rényi average lengths $L_T \to L_R$ as:*

$$\begin{aligned} L_R^* &= \frac{\ln_D \left[1 + \left(D^{\alpha-1} - 1 \right) H_D^{T \to R} \right]}{1-\alpha} \\ &= \frac{\ln_D \left[1 + \left(D^{\alpha-1} - 1 \right) \sum_j^{\Omega} p_{T \to R}(x_j) l_{T \to R}^*(x_j) \right]}{1-\alpha} \\ &= \frac{\ln_D \left[\sum_j^{\Omega} p_R^{\alpha}(x_j) \right]}{1-\alpha} \\ &= H_D^R(X), \end{aligned} \tag{4.54}$$

entitling a map between $l_T(x_j)$ and $l_R(x_j)$ by means of the relations

$$\begin{aligned} l_R^*(x_j) &= \frac{\ln_D \left[1 + \left(D^{\alpha-1} - 1 \right) l_{T \to R}^*(x_j) \right]}{1-\alpha}, \quad p_T \to p_R, \\ l_T^*(x_j) &= -\frac{D^{-(q-1) l_{R \to T}^*(x_j)} - 1}{D^{q-1} - 1}, \quad p_R \to p_T \,. \end{aligned} \tag{4.55}$$

It is easily proved that $L_R \geq H_D^R(X)$, which leads to the corresponding coding theorems, Refs.[11,78]. Despite we shall not further inspect such discussion, we observe from our previous exposition that to each information measure corresponds a specific optimal length, so that there exists a code whose mean length becomes optimal iff it attains the entropy. For that reason, one shall not expect two individual lengths $l(x_j)$ coincide if they come from a different measure of entropy, unless a good approximation or a particular case have been fetched.

Furthermore, the standard average length $\sum_j^\Omega p(x_j)l(x_j)$ can be seen as a linear combination of the individual lengths $l(x_j)$ weighted by the corresponding probabilities $p(x_j)$. Besides, in the standard formulation, the lengths obey $l^*(x_j) = -\ln_D p(x_j)$, meaning the codeword lengths will be long (short) for low (high) probabilities. Nonetheless, in general, the cost of using a codeword is not necessarily a linear function of its length, e.g. $H_D^\pm(X)$ or $H_D^R(X)$. Indeed, there are weights that privilege particular lengths such as $D\beta l_R(x_j)$ in (4.43), which encompasses an exponential law for the distribution of the lengths and, ultimately, raising the importance of those terms with low probabilities.

5

Superstatistical Measures of Complexity

In this chapter, a possible generalisation to the algorithmic theory of information is examined using the superstatistical formulation already studied in Chapter 2. The proposal is simple: It consists of relating the concept of generalised entropy to its corresponding complexity measure—also known as algorithmic entropy. Nevertheless, entropy as a measure of information must meet the stability criterion[50] to become eligible as a suitable tool for such purposes. Which reduces the whole universe of generalised entropies to a small set. To our knowledge, there are only two non-parametric entropies that generalise Shannon's entropy, while simultaneously fulfil the condition of stability, namely: the entropies (2.3) or (4.1). For this reason, we support the following discussion about these measures of entropy and the related measures of complexity, from now on identified as $K^{+}(X)$ and $K^{-}(X)$.

5.1 Algorithmic Superstatistics

In Chapter 2, we briefly discussed how the superstatistics framework[4] is employed to generate an infinite set of measures of entropy. Curiously, in the context of the algorithmic theory of information, one proceeds in an analogous way to obtain an infinite set of measures of complexity. However, to achieve these goals, certain concepts must be reworded.

To proceed, let us begin by summarising the basic ideas. The superstatistics approach addresses macroscopic systems outside equilibrium segmented into cells. These ones manifest asymptotic stationary states with a spatiotemporally fluctuating intensive quantity, which hardly varies over time, e.g. the inverse temperature β. There is a specific β assigned to each cell, so that β distributes according to a piecewise continuous probability density $f(\beta)$. Inside each region, β is approximately constant and, therefore, there exists a local equilibrium. However, at a global level, the system lies in a state out of equilibrium but enough isolated from external upheavals, thus behaving *slightly deviated from equilibrium*. When the entire fluctuations are averaged—inasmuch as the sum converges—an effective Boltzmann factor is obtained giving rise to generalised statistics. In addition, given the method relies on normalisable distributions $f(\beta)$, we are allowed to think of an infinite set of possible generalised statistics.

From the standpoint of computer science, we are conceiving of these cells as individual programs x that cast an outcome and halt, suggesting that the total system could be thought of as a universal computer U of general purpose, i.e. a Turing machine. In fact, in a real-world scenario, the tasks a computer surpasses are the results of a large collection of different recipes stored in it. In consequence, we say U is in the state x with probability $p(x)$. That is, each of these programs has a definite probability $p(x)$ of being taken at random by U.

Accordingly, if we denote with $|x|$ the length of the program x, hence the Boltzmann factor (2.4) becomes:

$$B_U(|x|) = \int_0^\infty \mathrm{d}\beta \, f_U(\beta) e^{-\beta|x|} \,, \quad \beta > 0 \,, \tag{5.1}$$

this is the effective Boltzmann factor related to a Turing machine U.

Remark 9. *If all cells share the same β, then the whole system will be considered as a single cell, in that case $f_U(\beta) = \delta(\beta - \beta_0)$, yielding the ordinary Boltzmann factor:*

$$B_U(|x|) = e^{-\beta|x|} \,.$$

Provided the generalised statistics must be normalisable over the whole domain of program lengths, the integral (partition function Z)

$$\int_0^\infty \mathrm{d}|x| \, B_U(|x|) \tag{5.2}$$

must converge.

Nevertheless, the partition function is oftentimes uncomputable. For example, substituting the ordinary Boltzmann factor $B_U(|x|) = e^{-\beta|x|}$ into (5.2), the integral will exist provided that $\beta \geq \ln 2$, even though, as shown in Ref.[74], the integral will be uncomputable and partially random for $\ln 2$. For our purposes we mus assume $f_U(\beta)$ is absolutely integrable.

Definition 9. *A function $f_U : \mathbb{R} \to \mathbb{R}$ is absolutely integrable when*

$$\int_{-\infty}^{\infty} \mathrm{d}t \, |f_U(t)| < \infty \, . \tag{5.3}$$

Let us now relate these concepts with algorithmic entropy. As we have previously discussed, the gist of the superstatistics approach is the effective Boltzmann factor $B_U(|x|)$ whose construction rests on a probability distribution well-nigh customised for a system of particular characteristics. In practice, this serves as a pivot to sprout general statistics in which the related fundamental quantities (entropy, free energy, etc.) must be rewritten in the frame of the new strategy.

The overall expression of entropy among superstatistics is particularly interesting. Its formal definition was already given in Chapter (2). We recall that we must consider only measures of entropy of the form (2.2).

In succession, assuming the integral (5.1) exists and can be computed, and since each program x lies in a set X of dimension Ω, we say the entropy is alternatively expressed as

$$-\sum_{x \in X} p(x) \Lambda_D(p(x)) \, ,$$

where $\Lambda_D(x)$ is an effective logarithm with base D (see Appendix B).

Some comments are in order. Notice the entropy is also interpreted as the expected value of $-\Lambda_D(p(x))$, $H_D(X) = \mathbb{E}[-\Lambda_D(p(x))]$. Although our notation varies somewhat from the previous chapter, it will be evident soon, it makes exposure easier in this context. As well, along this discussion all logarithms either effective or not are base 2, hence, from now on we shall drop D from our notation unless a different thing needs to be specified.

5.2 Superstatistical Measures of Complexity

To examine some of the aspects conveyed by the superstatistical approach to measures of complexity, henceforth we are to assume $p(x)$ belongs to a recursive probability distribution, i.e. distributions computable with a Turing machine U. Besides, we also assume U is running over a prefix-free domain X. Recall that a set of strings X is prefix-free if no string $x \in X$ is prefix of another string $x' \in X$.

The randomness of an object x can be measured in a number of ways[16,49]. There is a particular measure, that despite being uncomputable it is conceptually riveting. That is the *Kolmogorov complexity*[13,47,69] and is formally defined as follows.

Definition 10. *Kolmogorov complexity. Let U be a prefix-free Turing Machine, the complexity of the string y with respect to U is determined as*

$$K_U(y) = \min_x \{|x| : U(x) = y\} \ ,$$

that is the minimum possible length over all programs x with the halt property, whose outcome is y.

The quantity $K_U(y)$ has an intuitive but profound meaning. For a person describing the recipe for a certain dish to another individual, such that this one cannot construct a different interpretation of the directions for the correct realisation of such meal, then the number of bits in that communication constitutes an upper bound on $K_U(y)$.

Nonetheless, instead of considering the complexity associated with a program x, we are interested in the probability $p(x)$ associated with that program, i.e. we look for the way this (minimum-length) program is distributed over the domain of programs that achieve a specific outcome y. Thus rather than $K_U(y)$, from now on we refer to this quantity as

$$K(X) \equiv K_U(p(x)) \ .$$

Both $K_U(y)$ and $K(X)$ are measures of information. The first one strictly arises from combinatorial arguments, while the latter is a purely statistical measure of complexity that represents the average rate at which information is extracted from a combinatorial trial. Since the second measure of complexity depends on how the program x is distributed according to the law $p(x)$, there might be a number of ways in which such probability distribution can be maximised, depending upon the choice of entropy and maybe some inherent constraints that typically relate to system's attributes.

Theorem 3. *Let $p(x)$ be a recursive probability distribution. For any generic entropy $\sum_{x\in X} h(x)$ of the form (2.2), there exists a statistical measure of complexity $K(x)$ such that they satisfy the relation*

$$0 \leq \varphi^{-1}\left(\sum_{x\in X} p(x)\varphi(K(x))\right) - \sum_{x\in X} h(x) \leq \varphi(K(X)) \,, \tag{5.4}$$

where φ is the Nagumo-Kolmogorov function.

Proof. The Nagumo-Kolmogorov function provides an estimate of the cost of processing any information measure[12] as $\varphi(K(X))$, which is interpreted as the cost of treating specific rates of complexity.

Further, φ is not arbitrary but depends on the entropy functional that maximises the distribution $p(x)$, such that

$$0 \leq \varphi^{-1}\left(\sum_{x\in X} p(x)\varphi(K(x))\right) \,,$$

which is nothing but the average length (4.2).

Likewise, for any superstatistical entropy of the form (2.2), one obtains an effective coding theorem (cf. Chapter 4):

$$\sum_{x\in X} h(x) \leq \varphi^{-1}\left(\sum_{x\in X} p(x)\varphi(K(x))\right) , \tag{5.5}$$

where, as shown in Refs.[45,51], the complexity $K(x)$ and Chaitin's formulation of complexity[16] relates to each other through the formula

$$K(x) = -\Lambda(m(x)) + O(1) \,,$$

where

$$m(y) = \left\{\sum_{x\in X} 2^{-|x|} : U(x) = y\right\} \,, \tag{5.6}$$

in other words, a Turing machine U running over the programs X will print the output y with probability $m(y)$.

The summation over the whole set of programs in (5.6) is easily simplified on the grounds that there is only one program x in X whose outcome is y. To see this, imagine n programs $x_1, \ldots, x_n$, all having the same outcome y. Still, we are solely interested in the short one given that K regards the minimum possible description of y. Moreover, in the hypothetical case that X contains n *shortest* programs $x = x_1 = \cdots = x_n$, all printing the same output y, then $|x| > \ln n$.

Accordingly, Eq. (5.5) is equivalently expressed as:

$$\sum_{x\in X} h(x) \leq \varphi^{-1}\left(\sum_{x\in X} p(x)\varphi(|x|+O(1))\right) ,$$

reordering terms and invoking Theorem 2, we obtain the relation:

$$0 \leq \varphi^{-1}\left(\sum_{x\in X} p(x)\varphi(K(x))\right) - \sum_{x\in X} h(x) \leq \varphi(K(X)) ,$$

and we have the Theorem. □

This theorem states that entropy and complexity, as measures of information, are truly connected if the complexity is treated from a statistical viewpoint. Not only does the latter relation comprise the coding theorem formulated by Shannon[66], but it also suggests the entropy provides the minimum rate at which the complexity is expressed. Some of the consequences conveyed by Eq. (5.4), in terms of generalised entropies, shall be surveyed in the following section.

5.3 Superstatistical Measures of Complexity

As an application of Theorem 3, in this section we are to derive the effective measures of complexity associated with entropies (4.1):

$$H^{+}(X) = -\sum_{x\in X} p(x)\ln^{+}(p(x)) , \quad H^{-}(X) = -\sum_{x\in X} p(x)\ln^{-}(p(x)) , \tag{5.7}$$

whose series representations

$$\begin{aligned} H^{+}(X) &= -\sum_{x\in X}\sum_{k\in\mathbb{N}} \frac{[p(x)\ln p(x)]^{k}}{k!} , \\ H^{-}(X) &= -\sum_{x\in X}\sum_{k\in\mathbb{N}} (-1)^{k+1}\frac{[p(x)\ln p(x)]^{k}}{k!} , \end{aligned} \tag{5.8}$$

will simplify the following discussion. Besides, we are to focus on the nonequilibrium region associated with (5.7), since the limiting case $\Omega \to \infty$ coincides with the widely known Shannon's results.

Indeed, Theorems 1 and 2 state the minimum rate of data compression that are accomplished by a codification process is bounded from above and below by the measure

of entropy that characterises the statistics of the system involved. This result has been deeply studied in Chapter 4, still, it conforms a cornerstone for our current purposes, namely for the statement of the following theorem.

Lemma 2. *Let $p(x)$ be a recursive probability distribution. For a Nagumo-Kolmogorov cost function $\varphi(z) = z$, the measures of entropy $H^+(X)$ and $H^-(X)$ induce the existence of effective complexities $K^+(X)$ and $K^-(X)$ such that:*

$$0 \leq \sum_{x \in X} p(x) K^{\pm}(x) - H^{\pm}(X) \leq K^{\pm}(X) \, ,$$

where $K^+(X)$ and $K^-(X)$ are interpreted, respectively, as average lower and upper bounds on the statistical complexity K(X).

Proof. The inequality at the left implies

$$H^{\pm}(X) \leq \sum_{x \in X} p(x) K^{\pm}(x) \, ,$$

which is assured by Theorem 1 provided $K^{\pm}(x) \leq |x|^{\pm}$, with equality iff $|x|^{\pm} = |x|^{\pm *}$.

On the other hand, to prove the second inequality, suppose that

$$K^{\pm}(x) = c' |x|^{\pm *} \, , \quad c' \geq 1 \, ,$$

namely $K^{\pm}(x) + O(1) = |x|^{\pm *}$, then

$$c' \sum_{x \in X} p(x) |x|^{\pm *} + \sum_{x \in X} p(x) \ln^{\pm}(p(x)) \leq c' |X|^{\pm *} \, ,$$

regrouping terms on both sides, we get

$$c' \sum_{x \in X} \left(1 - p(x)^{-1}\right) p(x) \ln^{\pm}(p(x)) \leq \sum_{x \in X} p(x) \ln^{\pm}(p(x)) \, ,$$

since $p(x) \leq 1$, it follows that $1 - p(x)^{-1} \leq 0$, hence the inequality is true, while equality holds for $p(x) = 1$. That concludes the proof. □

It shall not escape attention $K^{\pm}(X)$ are not combinatorial measures of information but statistical measures of complexity and they shall not be interpreted as descriptive bounds on the Kolmogorov complexity as stated in Def. 10. Rather, what the measures $K^{\pm}(X)$ quantify is an average rate of complexity in agreement with that information estimated by entropies $H^{\pm}(X)$.

As an example, consider the probability distribution:

$$p(x) = \begin{cases} 0.y & \text{if } x = x_1 \\ 1 - 0.y & \text{if } x = x_2 \\ 0 & \text{otherwise,} \end{cases}$$

where y is the binary representation of a number $0.y$ between 0 and 1.

For the entropy measure $H^+(X)$ we have:

$$0 \leq (c' - 1)(-0.y \ln^+ 0.y - (1 - 0.y) \ln^+(1 - 0.y)) \leq c'(-\ln^+ 0.y - \ln^+(1 - 0.y)) \ ,$$

but $-x \ln^{\pm} x \leq -\ln^{\pm} x$, with equality if $x = 1$, then from the expression above we get

$$(c' - 1)(-\ln^+ 0.y - \ln^+(1 - 0.y)) \leq c'(-\ln^+ 0.y - \ln^+(1 - 0.y)) \ ,$$

analogously for $H^-(X)$.

5.3.1 Kullback-Leibler Divergence as a Measure of Complexity

At times, it may be worthwhile to quantify how different a given probability distribution is from another, e.g. a prior with respect to a trial distribution. In statistics this is performed by defining the entropy of the distribution p relative to another distribution q, to wit:

$$H(p \parallel q) = -\sum_{x \in X} p(x) \Lambda(p(x)) + \sum_{x \in X} p(x) \Lambda(q(x)) \ , \tag{5.9}$$

of course, if $p = q$ the outcome is zero. This expression stands for a generalisation of the Kullback-Leibler divergence[48].

As well, Eq. (5.9) is a measure of information gain[70,71]. This is fairly intuitive since q is known as the *prior* in the Bayesian probability theory and conveys all initial speculations about something before any observation is carried out. Even though, one should move with caution since the prior may haul redundancies, to give an idea, the special case $q(x) = 1/\dim(X) = 1/\Omega$ for all $x \in X$, leads again to the entropy up to a constant.

Nonetheless, as shown in Ref.[3], the prior distribution q would induce interesting results. Imagine the Turing machine U runs the program x, prints the outcome y

and halts, which is expressed in symbols as $U(x) = y$. In that regard, we reach a generalisation of Eq. (5.6):

$$q(y) = \left\{ \sum_{x \in X} \epsilon^{-\beta|x|} : U(x) = y \right\} , \tag{5.10}$$

which acts as a counterpart of $p \sim \epsilon^{-\beta|x|}$ over the set $\mathbb{N}$.

Indeed, using this prior, and following the same arguments given in Ref.[3], we are to show that one could think of (5.9) as a generalisation of a superstatistical algorithmic entropy.

Suppose we are particularly interested in those programs whose output is the string s, hence the auxiliar distribution that allows us to select that sort of programs is

$$p_y(s) = \begin{cases} 1 & \text{if } y = s \\ \\ 0 & \text{otherwise} , \end{cases} \tag{5.11}$$

then we compute the entropy of $p_y(s)$ relative to (5.10) to obtain

$$\begin{aligned} H(p_y \parallel q) &= -\sum_{l \in \mathbb{N}} p_y(l) \Lambda(p_y(l)) + \sum_{l \in \mathbb{N}} p_y(l) \Lambda \left(\sum_{x \in X} \epsilon^{-\beta|x|} : U(x) = y \right) \\ &= K_U(y) + \Lambda\left(Z\right) , \end{aligned} \tag{5.12}$$

where $Z = \sum_{x \in X} \epsilon^{-\beta|x|}$ is the partition function, and the algorithmic entropy reads

$$K_U(y) = -\Lambda \left(\sum_{x \in X} \epsilon^{-\beta|x|} : U(x) = y \right) , \tag{5.13}$$

suggesting relative entropy (5.12) would generalise algorithmic entropy in Def. 10, cf. Refs.[14,15,80].

Thereby, when the generalised logarithm and exponential functions in Eq. (5.12) reduce to the fundamental functions $\ln(x)$ and e^x, one simply obtains the algorithmic entropy (parallel to Shannon's entropy) reported by Baez[3], explicitly:

$$K_U(y) = -\ln \left(\sum_{x \in X} e^{-\beta|x|} : U(x) = y \right) ,$$

the special case $\beta = \ln 2$ yields $K_U(y) = |x| + O(1)$, which indicates the complexity of the shortest program $x \in X$ that prints y and halts, in agreement with Def. 10.

We now express Eqs. (5.12) and (5.13) by attending the structure of entropies $H^+(X)$ and $H^-(X)$. To simplify the calculations we shall resort to their series representations in Eq. (5.8).

Correspondingly, the entropy $H^+(X)$ of a distribution $p_y(s)$ relative to the prior $q(y)$ defined in Eq. (5.10), simply becomes:

$$\begin{aligned} H^+(p_y \parallel q) &= -\sum_{l\in\mathbb{N}}\sum_{k\in\mathbb{N}} \frac{1}{k!}\left[p_y(l)\ln p_y(l)\right]^k \\ &\qquad + \sum_{l\in\mathbb{N}}\sum_{k\in\mathbb{N}} \frac{1}{k!}\left[p_y(l)\ln\left(\sum_{x\in X} e^{-\beta|x|} : U(x)=y\right)\right]^k \\ &= -\sum_{k\in\mathbb{N}} \frac{1}{k!}\ln^k\left(\sum_{x\in X} e^{-\beta|x|} : U(x)=y\right) \\ &\qquad + \sum_{k\in\mathbb{N}} \frac{1}{k!}\ln^k\left(\sum_{x\in X} e^{-\beta|x|}\right) \\ &= K_U^+(y) + \sum_{k\in\mathbb{N}} \frac{1}{k!}\ln^k(Z)\ , \end{aligned} \tag{5.14}$$

where

$$K_U^+(y) = -\sum_{k\in\mathbb{N}} \frac{1}{k!}\ln^k\left(\sum_{x\in X} e^{-\beta|x|} : U(x)=y\right)\ , \tag{5.15}$$

is the effective algorithmic entropy that generalises the one given in Ref.[3] and consequently in Refs.[14,15,80].

Ditto, the entropy $H^-(X)$ of a distribution $p_y(s)$ relative to a prior $q(y)$, reads:

$$\begin{aligned} H^-(p_y \parallel q) &= -\sum_{l\in\mathbb{N}}\sum_{k\in\mathbb{N}} \frac{(-1)^{k+1}}{k!}\left[p_y(l)\ln p_y(l)\right]^k \\ &\qquad + \sum_{l\in\mathbb{N}}\sum_{k\in\mathbb{N}} \frac{(-1)^{k+1}}{k!}\left[p_y(l)\ln\left(\sum_{x\in X} e^{-\beta|x|} : U(x)=y\right)\right]^k \\ &= -\sum_{k\in\mathbb{N}} \frac{(-1)^{k+1}}{k!}\ln^k\left(\sum_{x\in X} e^{-\beta|x|} : U(x)=y\right) \\ &\qquad + \sum_{k\in\mathbb{N}} \frac{(-1)^{k+1}}{k!}\ln^k\left(\sum_{x\in X} e^{-\beta|x|}\right) \\ &= K_U^-(y) + \sum_{k\in\mathbb{N}} \frac{(-1)^{k+1}}{k!}\ln^k(Z)\ , \end{aligned} \tag{5.16}$$

where

$$K_U^-(y) = -\sum_{k\in\mathbb{N}} \frac{(-1)^{k+1}}{k!} \ln^k \left(\sum_{x\in X} e^{-\beta|x|} : U(x) = y \right) , \tag{5.17}$$

is the algorithmic entropy related to entropy $H^-(X)$.

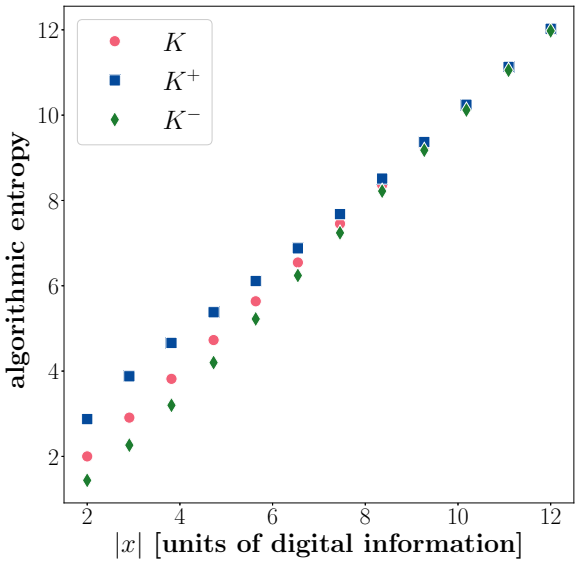


Figure 5.1: Measures of complexity or algorithmic entropies - Numerical comparison of effective algorithmic entropies, K, K^+ and K^- varying the data size $|x|$. The three measures converge as $|x|$ grows, while they differ from each other at extremely small size of information, below 8 bits. As of today, still there are applications depending on codes of such lengths. That is the case of the American Standard Code for Information Interchange (ASCII), which is an 8-bit code.

There is a compelling aspect we would like to highlight regarding the effective algorithmic entropies K^+ and K^-, derived from (5.14) and (5.16) respectively. Not only do they generalise the algorithmic entropy in Ref.[3], but they are both special cases of the relative entropies $H^+(p_y \parallel q)$ and $H^-(p_y \parallel q)$ while satisfying Lemma 2. Since these entropies enjoy of full stability, we think K^+ and K^- are both reasonable generalised measures of complexity.

We discussed the fact that the algorithmic entropies K^+ and K^- differ from the standard case K in a low-density program scheme, while the three measures coincide for enormous blocks of data. Although, in general, these functionals are uncomputable,

we still use a numerical trick to give an indication of their behaviour, see Fig. 5.1, where there is a region in which the algorithmic entropy K^- would account for a more economical description than the two other measures, yet as the program lengths $|x|$ increases the three measures tend to coincide. We cannot evaluate the precise impact of their real differences in describing complex objects. However, can those complex structures with a relatively small number of components enjoy a description different from that the standard formulation estimates?

6

Conclusions

Throughout this work, we studied the impact of entropies S_+ and S_- (or H^+ and H^-) on statistical mechanics, information theory and computer sciences. In all instances, they manifest differences with respect to the standard measures of entropy in physics and information sciences, as long as the number of available states is comparable to the number of particles or components of a given system, i.e. within a scheme of high probabilities. Whereas, in a regime of low probabilities, our proposal—which is free from ad hoc parameters—will retrieve the standard picture of thermostatistics.

Indeed, as we showed, the non-extensive entropies S_+ and S_- belong to the same asymptotic class that S_{BG}, which means they share the same properties in the thermodynamic limit.

Further, since S_+ and S_- are generalisations to the Boltzmann-Gibbs entropy, we proposed a possible extension to the entropic classification reported in Ref.[34], but in terms of the generators (2.11). It follows that, on the basis of entropies S_+ and S_-, we obtain a set of non-extensive entropies.

We compared S_+ and S_- with other proposals oftentimes referred in the literature, such as the Sharma-Mittal, Rényi and Tsallis entropies. However, these measures of entropy are unstable due to their dependence on free parameters. For example, we observed the Tsallis entropy, which is a special case of the generalised entropic forms in Eqs. (2.12)-(2.13), does not continuously resemble S_{BG} in the thermodynamic limit. This turns evident by applying the asymptotic laws (2.9)-(2.10) or by observing the leading contributions of the entropy in a microcanonic configuration.

6. CONCLUSIONS

Because of the stability S_+ and S_- enjoy, we were motivated to looking into further physical consequences conveyed by this kind of statistics; inasmuch as *stability* implies compatibility with observable phenomena[50]. Hence, we pursued an entropic derivation of Fokker-Planck equations, similar to the method discussed in Refs.[19,23].

We obtained two generalised Fokker-Planck equations outside the equilibrium. These models carry nonlinear terms that act as corrections to the mean-field Fokker-Planck equation. These corrections become unnecessary in a scenario where density ρ is very low, but at the mesoscopic level, nonlinear corrections could represent effective interactions between system components. Curiously, the first two nonlinear terms in the series representation of $F_{\pm}[\rho]$, in fact, correspond to the corrections introduced in some aggregation models studied in biology[77], although following a different approach.

We must remark these generalised Fokker-Planck equations are thermodynamically compatible with the usual Fokker-Planck equation, provided the equivalence class $(\gamma, c, d,) = \left(\frac{1}{2}, 1, 1\right)$ is common to S_+, S_- and S_{BG}.

Besides, observe from Eq. (3.19) the potential term is weighted by a function depending only on ρ. This function is univocally determined from the entropic form; an aspect that deserves attention since every generalisation to the Fokker-Planck equation must attend the suitable weight of the drift term to be consistent with the respective stochastic equation.

To compare Eqs. (3.19) with other models, we realised some numerical experiments. We found the weighted-drift terms in (3.19) are equivalent to effective potentials. In fact, we noted the drift terms

$$\chi\xi_+(p)\nabla\Phi \qquad \text{and} \qquad \chi\xi_-(p)\nabla\Phi$$

induce a stronger influence on the density ρ than the standard term

$$\chi\rho\nabla\Phi\ .$$

This behaviour allows us to claim that for systems owning few available microstates, the degree of heterogeneity is such that the interaction between its components produces non-negligible effective forces, which tend to vanish as the number of microstates increases substantially.

Likewise, the generalised diffusion models (3.23) portray anomalous diffusion. As mentioned earlier, we found our equations are directly comparable with those segregation models studied in Refs.[54,77]. We point out that, as far as Eqs. (3.23) are concerned, non-linear terms associated with interactions between organisms in biological contexts naturally appear as a consequence of generalised entropies S_+ and S_-.

We also reproduced numerically the general properties of the distributions that fit the transient diffusion model in Eq. (3.34), see Ref.[55]. In our proposal, this model is rewritten with regard to Eqs. (3.30). We observed interesting differences in the super-diffusion regime, mainly regarding the relaxation time. Under our model, differences in the relaxation time will occur due to non-equilibrium effects as well as intrinsic nonlinearities. It is also observed that equilibrium is restored after the entry into the super-diffusion regime, which is in line with the long-term recovery of the Boltzmann distribution.

In our investigation of the applications of non-extensive entropies and their effects, it was interesting for us to examine some optimisation processes in information theory. Above all, because entropy has a functional form similar to that used in the statistical formulation of thermodynamics. And also, because to some extent, the entropy interpretations are similar in both cases. The diligence of the entropies S_+ and S_- led us to introduce a non-extensive theory of information whose differences with the standard theory lie in regimes where the number of messages in a communication channel is comparable to the number of available states.

In a first step, using a variational method, we found the appropriate constraints to optimise the average lengths that a codeword achieves. These constraints, defined in Prop. 1, are not arbitrary provided the average lengths $L_\pm$ must equal the entropies $H_D^\pm(P)$ when they reach their optimal values.

Accordingly, we stated the *Generalised, Noiseless Coding Theorems* 1 and 2, that account for an effective data compression at different average rates, either in terms of $H_D^+(P)$ or $H_D^-(P)$. Therefore, the corresponding average lengths L_+ and L_- must be bounded from below and above in terms of the information measures themselves, which are directly associated with the optimal individual lengths $l_+(x_j)$ and $l_-(x_j)$.

Based on our findings, we noticed $H_D^+(P)$ might bring some novelty with respect to the data compression estimated through Shannon's entropy, in a low-density data pattern. We think such efficiency could be sustained in other regimes, if one pinpoints

a reasonable way to divide the system into sub-assemblies to outperform local data compression in each of them.

In addition, we studied the channel capacity of two simple generic models, a BSC and a BEC, regarding the entropies $H_D^+(P)$ and $H_D^-(P)$. In the BSC case, the channel capacity we obtained is higher than the estimated using Shannon's theory,

$$C_{\text{BSC}}^- \geq C_{\text{BSC}}^S \ ,$$

where the equality is attained at the probability values $p = \{0, 1/2, 1\}$.

Yet an interesting effect is observed in the capacities of a BEC, obtaining a twofold behaviour in each of the capacities C_{BEC}^+ and C_{BEC}^-, see Fig. 4.3. The swapping behaviour occurs due to the variation of *alpha*, which is the parameter representing the ratio of errors over the communication channel.

Our variational method can also be employed for other generalised measures of entropy. Without stating the corresponding coding theorems, we showed how the method works in the Tsallis and Rényi entropies.

Entropy is also the relationship between probabilistic and algorithm formulations of information theory. We explored one way to generalise the algorithmic formulation implementing the superstatistics framework. The consequences of such unification may lead to a differing understanding of the information processes.

The connection between the superstatistical framework and the Kolmogorov complexity comes immediately by following the relation (5.4), which is nothing but a generalisation of the statement formulated in Refs.[14,15,80], where Shannon's entropy is associated with the Kolmogorov complexity. Still, from our generalised statistical viewpoint, any measure of complexity constitutes an average rate of data compression that, in general, will differ from that appraised by the standard theory. This does not mean that all measures of complexity are truly feasible.

Note the effective Boltzmann factor (5.1) is calculated from a probability distribution that can be dependent on free parameters. While this parametric anatomy offers entropy flexibility sufficient to be potentially suitable for any circumstance, there is a trade-off with the functional's stability, which is a condition that must indistinctly be satisfied in order to qualify as an information measure and, therefore, as a measure of complexity.

However, we showed the stability H^+ and H^- own is inherited to K^+ and K^-, thus, they genuinely qualify as measures of complexity. To strengthen our arguments, we also showed that a generalised definition of relative entropy enables us to reconstruct the complexities K^+ and K^-, as specified in Eqs. (5.14) and (5.16). These results are equivalent to our preceding calculations, using Theorem 2.

As far as we have seen, there are entropy-induced effects produced by S_+ and S_- that are partially subtle in some probabilistic regimes. However, such modest differences led us to an effective theory of electron transport in super-diffuse environments that is consistent with what is observed in the laboratory. Whereas in terms of information, it led us to data compression that could be leveraged to other probabilistic orders. How much can *so little* mean?

Appendices

Appendix A

On the continuum representations of $S_{\pm}$

Let us begin with the known facts. The differential entropy

$$S_d = -\int \mathrm{d}x\, p(x) \ln p(x) \ ,$$

is not a limiting case of S_{BG} for $\Omega \to \infty$.

In fact, S_{BG} diverges from S_d if we outperform a successively fine space discretisation[63]. In other words, the growth rate given by entropy is in general not the same as the one calculated through the substitution

$$\sum_j \Delta x_j \to \int \mathrm{d}x \ .$$

In such a case, entropy may not remain invariant under coordinate transformations, even providing negative values. To overcome this difficulty, in the case of a limiting density of discrete points, a measure $m(x)$ must be introduced into the process in order to preserve invariance[40].

Hence, given an interval $[a, b]$ divided into Ω uniform subintervals

$$a = x_1 < x_2 < \cdots < x_\Omega = b \ ,$$

the limit $\Omega \to \infty$ reads as the density of points approaches a positive function $m(x)$. Furthermore, assuming the transition to the limit is sufficiently smooth so the adjacent differences between states is negligible, $(x_{j+1} - x_j) \sim 0$, then

$$\lim_{\Omega\to\infty} [(x_{j+1} - x_j)\Omega] = \frac{1}{m(x_j)} \ , \tag{A.1}$$

therefore, whenever the number of states Ω is large enough, the discrete probability distribution p_j and the continuous probability density $\rho(x_j)$ both relate to each other as

$$p_j = \rho(x_j)(x_{j+1} - x_j) = \rho(x_j)\Delta x_j \ , \tag{A.2}$$

leading to $p_j \to \rho(x_j)/\Omega m(x_j)$.

To investigate the continuous representation of $S_{\pm}$, we shall consider their series representations:

$$\begin{aligned} S_{\pm} &= \sum_{j=1}^{\Omega}\sum_{k=1}^{\infty} \frac{(\mp 1)^{k+1}}{k!}\left[p_j \ln\frac{1}{p_j}\right]^k \\ &= \sum_{j=1}^{\Omega}\sum_{k=1}^{\infty} \frac{(\mp 1)^{k+1}}{k!}\left[\rho(x_j)\Delta x_j \ln\frac{1}{p_j}\right]^k \\ &= \sum_{j=1}^{\Omega}\left[\rho(x_j)\Delta x_j \ln\frac{1}{p_j}\right] + \text{h.o.t.} \end{aligned} \tag{A.3}$$

where the higher-order terms (h.o.t.) equal

$$\sum_{j=1}^{\Omega}\sum_{k=1}^{\infty} \frac{(\mp 1)^{k+2}}{(k+1)!}\left[\rho(x_j)\Delta x_j \ln\frac{1}{p_j}\right]^{k+1} \ . \tag{A.4}$$

Notice the first term in Eq. (A.3) corresponds to the ordinary entropy, S_{BG}, whose continuum limit is given by Jaynes' formula[40], to wit:

$$\begin{aligned} \lim_{\Omega\to\infty} S_B &= \lim_{\Omega\to\infty}\left[\sum_{j=1}^{\Omega}\rho(x_j)\ln\left[\frac{m(x_j)}{\rho(x_j)}\right]\Delta x_j + \ln\Omega\right] \\ &= \int_a^b \mathrm{d}x\,\rho(x)\ln\left[\frac{m(x)}{\rho(x)}\right], \end{aligned} \tag{A.5}$$

where the diverging term, $\lim_{\Omega\to\infty}\ln\Omega$, has been discarded since it is the entropy change the observable to be measured[59].

We have to proceed in a similar fashion regarding (A.4), we get

$$\lim_{\Omega\to\infty}\text{h.o.t.} = \mp\lim_{\Omega\to\infty}\frac{1}{2!}\sum_{n=1}^{\Omega}\left(\rho(x_j)\Delta x_j\ln\left[\frac{\Omega m(x_j)}{\rho(x_j)}\right]\right)^2 \tag{A.6}$$

$$+\quad \lim_{\Omega\to\infty}\frac{1}{3!}\sum_{n=1}^{\Omega}\left(\rho(x_j)\Delta x_j\ln\left[\frac{\Omega m(x_j)}{\rho(x_j)}\right]\right)^3+\cdots$$

$$=\quad \mp\lim_{\Omega\to\infty}\frac{1}{2!}\sum_{n=1}^{\Omega}\sum_{n'=1}^{\Omega}\delta_{n,n'}\rho(x_j)\Delta x_j\ln\left[\frac{m(x_n)\Omega}{\rho(x_j)}\right]\rho(x_{n'})\Delta x_{n'}\ln\left[\frac{m(x_{n'})\Omega}{\rho(x_{n'})}\right]$$

$$+\quad \lim_{\Omega\to\infty}\frac{1}{3!}\sum_{n=1}^{\Omega}\sum_{n'=1}^{\Omega}\sum_{n''=1}^{\Omega}\delta_{n,n'}\delta_{n,n''}\rho(x_j)\Delta x_j\ln\left[\frac{m(x_n)\Omega}{\rho(x_j)}\right]\rho(x_{n'})\Delta x_{n'}\ln\left[\frac{m(x_{n'})\Omega}{\rho(x_{n'})}\right]$$

$$\times\quad \rho(x_{n''})\Delta x_{n''}\ln\left[\frac{m(x_{n''})\Omega}{\rho(x_{n''})}\right]+\cdots$$

$$=\quad \mp\frac{1}{2!}\int_a^b\mathrm{d}x\int_a^b\mathrm{d}x'\delta(x-x')\rho(x)\ln\left[\frac{m(x)\Omega}{\rho(x)}\right]\rho(x')\ln\left[\frac{m(x')\Omega}{\rho(x')}\right]\frac{1}{m(x)\Omega}$$

$$+\frac{1}{3!}\int_a^b\mathrm{d}x\int_a^b\mathrm{d}x'\int_a^b\mathrm{d}x''\delta(x-x')\delta(x-x'')\rho(x)\ln\left[\frac{m(x)\Omega}{\rho(x)}\right]\rho(x')\ln\left[\frac{m(x')\Omega}{\rho(x')}\right]$$

$$\times\quad \rho(x'')\ln\left[\frac{m(x'')\Omega}{\rho(x'')}\right]\frac{1}{m(x)^2\Omega^2}+\cdots\,.$$

In the last step we have applied the formula

$$\hat{\delta}_{j,j'} = m(x_j)\Omega\delta_{j,j'}\to\delta(x-x')\,, \tag{A.7}$$

with

$$\sum_j\hat{\delta}_{j,j'}\Delta x_j = 1\,. \tag{A.8}$$

Hence, we found that transition of $S_\pm$ from discrete to continuous leads to

$$S_\pm^{(c)} = \lim_{\Omega\to\infty}(S_B+\text{h.o.t.})\,,$$

however, provided the contribution of higher-order terms to the continuum is negligible, let us conclude that Jayne's formula corresponds to the continuous representation of $S_\pm$.

A. ON THE CONTINUUM REPRESENTATIONS OF $S_{\pm}$

Appendix B

Generalised Logarithms and Exponentials

Let $\epsilon : \mathbb{R} \to \mathbb{R}$ be a stretched exponential. Those functions $\Lambda : \mathbb{R} \to \mathbb{R}$ satisfying the conditions $\Lambda(1) = 0$ and $\Lambda'(1) = 1$, such that $\Lambda(\epsilon^x) = \epsilon^{\Lambda(x)} = x$ are called generalised (or effective) logarithms. Their series representation can usually be put into terms of the fundamental logarithm functions ln or log. For a more detailed discussion than the one presented here, see Ref.[35]. We limit to present the basic structure of the effective logarithms $\ln_+$ and $\ln_-$, and their inverses.

We have the functions

$$\begin{aligned} \ln^+(x) &\equiv -\frac{1-x^x}{x} \\ \ln^-(x) &\equiv -\frac{x^{-x}-1}{x}, \end{aligned} \tag{B.1}$$

for $x \in [0,1]$, otherwise the functions become undefined. From such definitions it becomes evident that the functions $\ln^{(\pm)}$ do not fulfil the three laws of logarithms. Yet they can be expanded in series as

$$\ln^+(x) = \ln x + \frac{1}{2!}x\ln^2 x + \frac{1}{3!}x^2\ln^3 x + \frac{1}{4!}x^3\ln^4 x + \cdots, \tag{B.2}$$

and

$$\ln^-(x) = \ln x - \frac{1}{2!}x\ln^2 x + \frac{1}{3!}x^2\ln^3 x - \frac{1}{4!}x^3\ln^4 x + \cdots, \tag{B.3}$$

note that the first term is in both cases leads the series, while higher order terms become subdominant as $x \to 0$. This peculiar flexibility grants to entropies (5.7)

the simultaneous character of accounting for non-equilibrium phenomena in the low-probability regime, while preserving a well defined thermodynamical limit.

The corresponding stretched exponentials of (B.1) do not posses a closed form, in this case we make use of a numerical representation. These functions have been constructed as

$$\exp^{\pm}(x) \equiv \exp(-x) \sum_{j=0}^{\infty} a_{\pm}(j) x^j, \quad a_{\pm}(j) \in \mathbb{R}, \tag{B.4}$$

the first nine coefficients $a_{\pm}(j)$ are given in Table B.1.

	$a_+(j)$	$a_-(j)$
$j = 8$	-0.000157095	0.000105402
$j = 7$	0.00373467	-0.00211934
$j = 6$	-0.0362676	0.0166679
$j = 5$	0.186358	-0.0675544
$j = 4$	-0.546751	0.16867
$j = 3$	0.905157	-0.317048
$j = 2$	-0.709322	0.3725
$j = 1$	0.0228963	0.0147449
$j = 0$	1	1

Table B.1: $a_{\pm}(j)$ coefficients.

References

[1] Sumiyoshi Abe. A note on the q-deformation-theoretic aspect of the generalized entropies in nonextensive physics. *Physics Letters A*, 224(6):326 – 330, 1997. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(96)00832-8. URL http://www.sciencedirect.com/science/article/pii/S0375960196008328. 2

[2] M.B. Baer. Source coding for quasiarithmetic penalties. *IEEE Transactions on Information Theory*, 52(10):4380–4393, Oct 2006. ISSN 1557-9654. doi: 10.1109/TIT.2006.881728. 45

[3] J. Baez and M. Stay. Algorithmic thermodynamics. *Mathematical Structures in Computer Science*, 22(5):771–787, 2012. 70, 71, 72, 73

[4] C. Beck and E.G.D. Cohen. Superstatistics. *Physica A: Statistical Mechanics and its Applications*, 322:267 – 275, 2003. ISSN 0378-4371. doi: https://doi.org/10.1016/S0378-4371(03)00019-0. URL http://www.sciencedirect.com/science/article/pii/S0378437103000190. 2, 10, 63

[5] B.M. Boghosian. Thermodynamic description of the relaxation of two-dimensional turbulence using Tsallis statistics. *Phys. Rev. E*, 53:4754–4763, May 1996. doi: 10.1103/PhysRevE.53.4754. URL https://link.aps.org/doi/10.1103/PhysRevE.53.4754. 20

[6] L. Boltzmann. *Lectures on Gas Theory*. Dover Books on Physics. Dover Publications, 1995. ISBN 9780486684550. URL https://books.google.com.mx/books?id=-I7QzCXnstEC. 1

[7] J.P. Bouchaud and A. Georges. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Physics Re-*

ports, 195(4):127 – 293, 1990. ISSN 0370-1573. doi: https://doi.org/10.1016/0370-1573(90)90099-N. URL http://www.sciencedirect.com/science/article/pii/037015739090099N. 24

[8] N. Cabo and O. Obregón. Exploring the gauge/gravity duality of a generalized von Neumann entropy. *Eur. Phys. J. P.*, 133(2):55, 2018. ISSN 2190-5444. doi: 10.1140/epjp/i2018-11883-5. 2, 12

[9] N. Cabo, J. Fuentes, and O. Obregón. Generalised asymptotic classes for additive and non-additive entropies. *EPL (Europhysics Letters)*, 128(6):60004, feb 2020. doi: 10.1209/0295-5075/128/60004. URL https://doi.org/10.1209/0295-5075/128/60004. 3, 99

[10] C. Callan and F. Wilczek. On geometric entropy. *Physics Letters B*, 333(1):55 – 61, 1994. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(94)91007-3. URL http://www.sciencedirect.com/science/article/pii/0370269394910073. 8

[11] L.L. Campbell. A coding theorem and Rényi's entropy. *Information and Control*, 8(4):423–425, August 1965. 57, 60, 61, 62

[12] L.L. Campbell. Definition of entropy by means of a coding problem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 6(2):113–118, Jun 1966. ISSN 1432-2064. doi: 10.1007/BF00537132. URL https://doi.org/10.1007/BF00537132. 45, 67

[13] G.J. Chaitin. On the simplicity and speed of programs for computing infinite sets of natural numbers. *J. ACM*, 16(3):407–422, July 1969. ISSN 0004-5411. doi: 10.1145/321526.321530. URL https://doi.org/10.1145/321526.321530. 66

[14] G.J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, July 1975. ISSN 0004-5411. doi: 10.1145/321892.321894. URL https://doi.org/10.1145/321892.321894. 71, 72, 78

[15] G.J. Chaitin. Algorithmic entropy of sets. *Computers and Mathematics with Applications*, 2(3):233 – 245, 1976. ISSN 0898-1221. doi: https://doi.org/10.1016/0898-1221(76)90016-X. URL http://www.sciencedirect.com/science/article/pii/089812217690016X. 71, 72, 78

[16] G.J. Chaitin. *Exploring Randomness*. Springer-Verlag London, 1 edition, 2001. ISBN 978-1-4471-0307-3. doi: 10.1007/978-1-4471-0307-3. 66, 67

[17] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89, Jan 1943. doi: 10.1103/RevModPhys.15.1. URL https://link.aps.org/doi/10.1103/RevModPhys.15.1. 25

[18] P.-H. Chavanis. Nonlinear mean-field Fokker-Planck equations and their applications in physics, astrophysics and biology. *Comptes Rendus Physique*, 7 (3):318 – 330, 2006. ISSN 1631-0705. doi: https://doi.org/10.1016/j.crhy.2006.01.004. URL http://www.sciencedirect.com/science/article/pii/S1631070506000284. Statistical mechanics of non-extensive systems. 30

[19] P.-H. Chavanis. Nonlinear mean-field Fokker-Planck equations. Application to the chemotaxis of biological populations. *The European Physical Journal B*, 62(2): 179–208, Mar 2008. ISSN 1434-6036. doi: 10.1140/epjb/e2008-00142-9. URL https://doi.org/10.1140/epjb/e2008-00142-9. 24, 26, 30, 76

[20] R. Clausius. *The Mechanical Theory of Heat: With its Applications to the Steam-Engine and to the Physical Properties of Bodies*. J. Van Voorst, 1867. 1

[21] D.S. Cohen and J.D. Murray. A generalized diffusion model for growth and dispersal in a population. *J. Math. Biology*, 12:237, 1981. 31

[22] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2012. ISBN 978-1-118-58577-1. 46, 53

[23] D. Czégel, S.G. Balogh, P. Pollner, and G. Palla. Phase space volume scaling of generalized entropies and anomalous diffusion scaling governed by corresponding non-linear Fokker-Planck equations. *Scientific Reports*, 8(1883), 2018. 24, 26, 76

[24] A. Deppman, E. Megías, and D. P. Menezes. Fractal structure of Yang-Mills fields. In *8th International Conference on New Frontiers in Physics*, 2 2020. 2

[25] A.A. Dubkov, B. Spagnolo, and V.V. Uchaikin. Lévy flight superdiffusion: An introduction. *International Journal of Bifurcation and Chaos*, 18(09):2649–2672, 2008. doi: 10.1142/S0218127408021877. URL https://doi.org/10.1142/S0218127408021877. 24

[26] A. Feinstein. *Foundations of information theory*. McGraw-Hill electrical and electronic engineering series. McGraw Hill: New York, 1958. URL https://books.google.com.mx/books?id=BtQ-AAAAIAAJ. 45, 55

[27] A.D. Fokker. Die mittlere energie rotierender elektrischer dipole im strahlungsfeld. *Annalen der Physik*, 348(5):810–820, 1914. doi: https://doi.org/10.1002/andp.19143480507. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19143480507. 24

[28] F. Franchini, A.R. Its, and V.E. Korepin. Renyi entropy of the XY spin chain. *Journal of Physics A: Mathematical and Theoretical*, 41(2):025302, dec 2007. doi: 10.1088/1751-8113/41/2/025302. URL https://iopscience.iop.org/article/10.1088/1751-8113/41/2/025302. 20

[29] J. Fuentes and O. Obregón. On how generalised entropies without parameters impact information optimisation processes. *arXiv:2103.11759 (under review)*, 2021. URL https://arxiv.org/abs/2103.11759. 4, 99

[30] J. Fuentes and O. Obregón. A superstatistical formulation of complexity measures. *arXiv:2101.09147 (under review)*, 2021. URL https://arxiv.org/abs/2101.09147. 5

[31] J. Fuentes, J. L. López, and O. Obregón. Generalized Fokker-Planck equations derived from nonextensive entropies asymptotically equivalent to Boltzmann-Gibbs. *Phys. Rev. E*, 102:012118, Jul 2020. doi: 10.1103/PhysRevE.102.012118. URL https://link.aps.org/doi/10.1103/PhysRevE.102.012118. 4, 99

[32] C.N. Gagatsos, A.I. Karanikas, and G. Kordas. Mutual information and bose-einstein condensation. *Open Systems & Information Dynamics*, 20(02):1350008, 2013. doi: 10.1142/S123016121350008X. URL https://doi.org/10.1142/S123016121350008X. 8

[33] A. Gil-Villegas, O. Obregón, and J. Torres-Arenas. Computer simulation of effective potentials for generalized Boltzmann-Gibbs statistics. *Journal of Molecular Liquids*, 248:364 – 369, 2017. ISSN 0167-7322. doi: https://doi.org/10.1016/j.molliq.2017.10.027. URL http://www.sciencedirect.com/science/article/pii/S0167732217328738. 11

[34] R. Hanel and S. Thurner. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *Eur. Phys. Lett.*, 93:20006, 2011. 12, 14, 15, 75

[35] R. Hanel, S. Thurner, and M. Gell-Mann. Generalized entropies and logarithms and their duality relations. *Proceedings of the National Academy of Sciences*, 109 (47):19151–19154, 2012. ISSN 0027-8424. doi: 10.1073/pnas.1216885109. URL https://www.pnas.org/content/109/47/19151. 87

[36] R. Hanel, S. Thurner, and M. Gell-Mann. Generalized entropies and logarithms and their duality relations. *Proceedings of the National Academy of Sciences*, 109 (47):19151–19154, 2012. ISSN 0027-8424. doi: 10.1073/pnas.1216885109. URL https://www.pnas.org/content/109/47/19151. 25

[37] J. Havrda and F. Charvát. Quantification method of classification processes. *Kybernetika*, 3:30–35, 1967. 20, 57

[38] T. Ichibayashi and K. Tanimura. Ultrafast carrier relaxation in si studied by time-resolved two-photon photoemission spectroscopy: Intravalley scattering and energy relaxation of hot electrons. *Phys. Rev. Lett.*, 102:087403, Feb 2009. doi: 10.1103/PhysRevLett.102.087403. URL https://link.aps.org/doi/10.1103/PhysRevLett.102.087403. 39

[39] V.M. Ilic, I. Djordjevic, and F. Küppers. On the Daróczy-Tsallis capacities of discrete channels. In *2nd International Electronic Conference on Entropy and Its Applications session Information Theory*, volume 2. SciForum, MDPI, 2015. doi: DOI:10.3390/ECEA-2-B004. 55

[40] E.T. Jaynes. Information theory and statistical mechanics. In K.W. Ford, editor, *Statistical Physics*, volume 3, page 181. Brandeis Lectures in Theoretical Physics, W.A. Benjamin, Inc., 1963. 83, 84

[41] D. Kabat. Black hole entropy and entropy of entanglement. *Nuclear Physics B*, 453(1):281 – 299, 1995. ISSN 0550-3213. doi: https://doi.org/10.1016/0550-3213(95)00443-V. URL http://www.sciencedirect.com/science/article/pii/055032139500443V. 8

[42] D. Kabat and M.J. Strassler. A comment on entropy and area. *Physics Letters B*, 329(1):46 – 52, 1994. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(94)90515-0. URL http://www.sciencedirect.com/science/article/pii/0370269394905150. 8

[43] G. Kaniadakis. Non-linear kinetics underlying generalized statistics. *Physica A: Statistical Mechanics and its Applications*, 296(3):405 – 425, 2001. ISSN 0378-4371. doi: https://doi.org/10.1016/S0378-4371(01)00184-4. URL http://www.sciencedirect.com/science/article/pii/S0378437101001844. 2

[44] A.I. Khinchin. *Mathematical Foundation of Information Theory*. Dover Publications, New York., 1957. 8

[45] W. Kirchherr, M. Li, and P. Vitányi. The miraculous universal distribution. *The Mathematical Intelligencer*, 19(4):7–15, 1997. doi: 10.1007/BF03024407. URL https://doi.org/10.1007/BF03024407. 67

[46] J.G. Kirkwood. The statistical mechanical theory of transport processes i. general theory. *The Journal of Chemical Physics*, 14(3):180–201, 1946. doi: 10.1063/1.1724117. URL https://doi.org/10.1063/1.1724117. 25

[47] A.N. Kolmogorov. Three approaches to the definition of the concept quantity of information. *Probl. Peredachi Inf.*, 1(1):3–11, 1965. URL http://mi.mathnet.ru/ppi68. 66

[48] S. Kullback and R.A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951. doi: 10.1214/aoms/1177729694. URL https://doi.org/10.1214/aoms/1177729694. 70

[49] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22(1):75–81, 1976. 66

[50] B. Lesche. Instabilities of Rényi entropy. *J. Stat. Phys.*, 27(2):419–422, 1982. 17, 63, 76

[51] L.A. Levin. Universal sequential search problems. *Problems Inform. Transmission*, 9(3):115–116, 1973. 67

[52] J. A. S. Lima and A. Deppman. Tsallis meets Boltzmann: q-index for a finite ideal gas and its thermodynamic limit. *Phys. Rev. E*, 101:040102(R), Apr 2020. doi: 10.1103/PhysRevE.101.040102. URL https://link.aps.org/doi/10.1103/PhysRevE.101.040102. 2

[53] W.-T. Lin and C.-L. Ho. Similarity solutions of the fokker–planck equation with time-dependent coefficients. *Annals of Physics*, 327(2):386 – 397, 2012. ISSN 0003-4916. doi: https://doi.org/10.1016/j.aop.2011.11.004. URL http://www.sciencedirect.com/science/article/pii/S0003491611001771. 24

[54] V. Mendez, D. Campos, and F. Bartumeus. *Stochastic Foundations in Movement Ecology: Anomalous Diffusion, Front Propagation and Random Searches*. Springer Complexity, 1 edition, 2014. ISBN 978-3-642-39009-8. doi: 10.1007/978-3-642-39010-4. 37, 77

[55] E. Najafi, V. Ivanov, A. Zewail, and M. Bernardi. Super-diffusion of excited carriers in semiconductors. *Nature Communications*, 8:15177 EP –, 05 2017. URL https://doi.org/10.1038/ncomms15177. 39, 40, 77

[56] O. Obregón. Superstatistics and gravitation. *Entropy*, 12:2067, 2010. 2, 10

[57] O. Obregón. Generalized information and entanglement entropy, gravitation and holography. *Int. J. Mod. Phys. A*, 30(16):1530039, 2015. 11

[58] O. Obregón, J. L. López, and M. Ortega-Cruz. On quantum superstatistics and the critical behavior of nonextensive ideal Bose gases. *Entropy*, 20(10):773, 2018. ISSN 1099-4300. doi: 10.3390/e20100773. URL https://www.mdpi.com/1099-4300/20/10/773. 2, 11

[59] T. Oikonomou and G. B. Bagci. Route from discreteness to the continuum for the tsallis q-entropy. *Phys. Rev. E*, 97:012104, Jan 2018. doi: 10.1103/PhysRevE.97.012104. URL https://link.aps.org/doi/10.1103/PhysRevE.97.012104. 84

[60] O. Parkash and P. K. Sharma. Noiseless coding theorems corresponding to fuzzy entropies. In *10th IEEE International Conference on Fuzzy Systems. (Cat. No.01CH37297)*, volume 1, pages 176–179 vol.1, Dec 2001. doi: 10.1109/FUZZ.2001.1007276. 57, 59

[61] M. Planck. Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie. In *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*, volume 24, pages 324–341. Köninglich Preussischen, 1917. 24

[62] A.R. Plastino and A. Plastino. Non-extensive statistical mechanics and generalized fokker-planck equation. *Physica A: Statistical Mechanics and its Applications*, 222(1):347 – 354, 1995. ISSN 0378-4371. doi: https://doi.org/10.1016/0378-4371(95)00211-1. URL http://www.sciencedirect.com/science/article/pii/0378437195002111. 2

[63] A. Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(1):193–215, Mar 1959. ISSN 1588-2632. doi: 10.1007/BF02063299. URL https://doi.org/10.1007/BF02063299. 83

[64] A. Rényi. On measures of information and entropy. *Proc. Fourth Berkeley Symp. Math. Stats. Prob.*, pages 547–561, 1961. 18, 59

[65] S.A. Rice and A.R. Allnatt. Approximate theory of transport in simple dense fluid mixtures. *The Journal of Chemical Physics*, 34(2):409–420, 1961. doi: 10.1063/1.1700966. URL https://doi.org/10.1063/1.1700966. 25

[66] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948. ISSN 0005-8580. doi: 10.1002/j.1538-7305.1948.tb01338.x. 1, 43, 46, 68

[67] B.D. Sharma and D.P. Mittal. New nonadditive measures of entropy for discrete probability distributions. *Math. Sci. (Soc. Math. Sci., Delhi, India)*, 10:28–40, 1975. 18

[68] F. Shibata, Y. Takahashi, and N. Hashitsume. A generalized stochastic Liouville equation. Non-Markovian versus memoryless master equations. *Journal of Statistical Physics*, 17(4):171–187, Oct 1977. ISSN 1572-9613. doi: 10.1007/BF01040100. URL https://doi.org/10.1007/BF01040100. 2

[69] R.J. Solomonoff. A preliminary report on a general theory of inductive inference. In *Technical Report ZTB-138, Zator Co. and Air Force Office of Scientific Research*. Cambridge, Mass., 1960. URL http://world.std.com/~rjs/publications/z138.pdf. 66

[70] R.J. Solomonoff. A formal theory of inductive inference. part i. *Information and Control*, 7(1):1 – 22, 1964. ISSN 0019-9958. doi: https://doi.org/10.1016/S0019-9958(64)90223-2. URL http://www.sciencedirect.com/science/article/pii/S0019995864902232. 70

[71] R.J. Solomonoff. A formal theory of inductive inference. part ii. *Information and Control*, 7(2):224 – 254, 1964. ISSN 0019-9958. doi: https://doi.org/10.1016/S0019-9958(64)90131-7. URL http://www.sciencedirect.com/science/article/pii/S0019995864901317. 70

[72] S. Sootla, D. O. Theis, and R. Vicente. Analyzing information distribution in complex systems. *Entropy*, 19(12), 2017. ISSN 1099-4300. doi: 10.3390/e19120636. URL https://www.mdpi.com/1099-4300/19/12/636. 2

[73] B. Streetman and S. Banerjee. *Solid State Electronic Devices*. Cloth, 5th edition, 2006. ISBN 9780131497269. 39

[74] K. Tadaki. A generalization of Chaitin's halting probability ω and halting self-similar sets. *Hokkaido Math. J.*, 31(1):219–253, 02 2002. doi: 10.14492/hokmj/1350911778. URL https://doi.org/10.14492/hokmj/1350911778. 65

[75] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.*, 52:479, 1988. 18, 57

[76] C. Tsallis and A.M.C. Souza. Constructing a statistical mechanics for beck-cohen superstatistics. *Physical Review E*, 67(2):026106, 2003. 11

[77] P. Turchin. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Weimar and Now; 13. Sinauer, 1998. ISBN 9780878938476. URL https://books.google.com.mx/books?id=ZbdmQgAACAAJ. 32, 36, 76, 77

[78] H. Tyagi. Coding theorems using rényi information measures. In *2017 Twenty-third National Conference on Communications (NCC)*, pages 1–6, March 2017. doi: 10.1109/NCC.2017.8077037. 62

[79] J. Von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer-Verlag Berlin Heidelberg, 2 edition, 1932. ISBN 978-3-540-59207-5. doi: 10.1007/978-3-642-61409-5. 1

[80] A.K. Zvonkin and L.A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83, 1970. 71, 72, 78

Declaration

Under protest of telling the truth, I declare this dissertation is the result of my own research and that of my co-authors and that it has not been subject to any prior evaluation to obtain an academic degree.

The contents of this work, such as figures, tables, theorems, lemmas and their proofs are of our intellectual authorship and were reported in peer-review publications in Refs.[9,29–31].

The thesis work was conducted from 2018 to 2021 under the supervision of Dr. Octavio Obregón at the Division of Sciences and Engineering of the University of Guanajuato.

1 March 2021
León, Guanajuato, México

León, Guanajuato a 24 de Junio de 2021
Asunto: Carta de revisión de tesis de doctorado de Jesús Fuentes Aguilar

DR. DAVID YVES GHISLAIN DELEPINE
Director de la División de Ciencias e Ingenierías, Campus León
Universidad de Guanajuato
P R E S E N T E

Estimado Dr. Delepine,

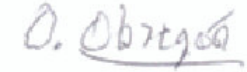
Como miembro del comité de sinodales del estudiante de Doctorado en Física, Jesús Fuentes Aguilar, le informo que he revisado su tesis titulada “On Non-extensive Entropies: With Applications in Stochastic Dynamics and Information Theory“, misma que Jesús desarrolló bajo mi supervisión.

Como resultado de la tesis, se han publicado dos artículos en revistas con arbitraje internacional, estos son “*Generalised asymptotic classes for additive and non-additive entropies*” EPL 128, 60004 (2019) y “*Generalized Fokker-Planck equations derived from nonextensive entropies asymptotically equivalent to Boltzmann-Gibbs*” Phys. Rev. E 102, 012118 (2020). Además hay otros dos trabajos en proceso, titulados “*Optimisation of information processes using non-extensive entropies without parameters*” y “*A Superstatistical Formulation of Complexity Measures*”.

Considero que dicho trabajo cumple con los requisitos necesarios para ser defendido públicamente y completar el programa de doctorado.

Sin otro particular por el momento, le envío un cordial saludo.

ATENTAMENTE



Dr. Octavio Obregón

UNIVERSIDAD DE GUANAJUATO

León Gto. a 21 de mayo de 2021
Asunto: Carta de revisión de tesis de Jesús Fuentes Aguilar

Dr. David Delepine
Director de la División de Ciencias e Ingenierías, Campus León
Universidad de Guanajuato
PRESENTE

Estimado Dr. Delepine,

En mi calidad miembro del comité de sinodales del alumno de Doctorado en Física Jesús Fuentes Aguilar, por este medio informo a usted que he revisado su tesis titulada *"On Non-Extensive Entropies: With Applications in Stochastic Dynamics and Information theory"* que desarrolló Jesús con el fin de obtener el grado de Doctor en Física.

El trabajo de Jesús Fuentes posee el contenido, originalidad y relevancia necesaria como trabajo de investigación y Jesús ha tomado en cuenta mis observaciones en la versión final de la tesis. Considero que dicho trabajo está listo para ser defendido públicamente.

Sin más por el momento, me despido con un cordial saludo.

ATENTAMENTE
"LA VERDAD OS HARÁ LIBRES"

Dr. José Luis López Picón
Departamento de Física
División de Ciencias e Ingenierías, Campus León

Alberto Diez Tejedor
Profesor Asociado C
Departamento de Física
DCI-León

León, Guanajuato; a 8 de julio de 2021

Dr. David Yves Ghislain Delepine
Director de la División de Ciencias e Ingenierías
Campus León, Universidad de Guanajuato
P R E S E N T E

Estimado Dr. Delepine,

Por este medio, me permito informarle que he leído y revisado la tesis titulada **"On Non-Extensive Entropies: With Applications in Stochastic Dynamics and Information Theory"**, que realizó el estudiante **Jesús Fuentes Aguilar** como requisito para obtener el grado de Doctor en Física.

Considero que el trabajo de tesis realizado por Jesús reúne los requisitos necesarios de calidad e interés académico para que sea defendido en un examen de grado, razón por la cual extiendo mi aval para que así se proceda.

Sin más que agregar, agradezco su atención y aprovecho la ocasión para enviarle un cordial saludo.

ATENTAMENTE
"LA VERDAD OS HARÁ LIBRES"

Dr. Alberto Diez Tejedor
Departamento de Física
DCI, Campus León

UNIVERSIDAD DE GUANAJUATO

León, Gto., a 30 de junio de 2021
Asunto: **Revisión de Tesis**

DR. DAVID YVES GHISLAIN DELEPINE
DIRECTOR
DIVISIÓN DE CIENCIAS E INGENIERIAS
CL -UNIVERSIDAD DE GUANAJUATO

A través de la presente constato que he revisado la tesis del C. **Jesús Fuentes Aguilar** con el fin de obtener el grado de Doctorado en Física. El trabajo de tesis se titulado **"On Non-Extensive Entropies: With Applications in Stochastic Dynamics and Information Theory"**. En este estudio realizado por Jesús desarrolla un estudio de las propiedades y aplicaciones de las teorías de una variante de entropía no extensiva, en la cual demuestra aplicaciones y sus efectos en complejidad cuántica y en compresión de información, así como en dinámica aplicada a sistemas biológicos. El trabajo de titulación satisface con la completez y solidez de un proyecto de titulación a nivel doctorado y además cuenta con varias publicaciones en revistas indexadas. Jesús ha realizado las correcciones recomendadas al documento de la tesis. Además, he cuestionado a Jesús en los temas relacionados a su trabajo de tesis, demostrando su dominio en los temas abordados en su trabajo de tesis. Por lo que considero que puede proceder con los trámites para la disertación de tesis.

Sin más por el momento le envío saludos cordiales.

Atentamente

Dr. Carlos Herman Wiechers Medina
Profesor-Investigador

Tel. +52 (477) 7885100 Ext. 8467
Cel. +52 (477) 1080605
e-mail 1: carherwm@fisica.ugto.mx
e-mail 2: ch.wiechers@ugto.mx

Centro de Investigación y de Estudios Avanzados del IPN
Apartado Postal 14-740
México 07000, D.F.

Departamento de Física
Tel. 57 47 38 00 ext 6117
Fax 57 47 70 96

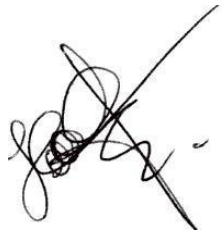
5 de Julio, 2021

Dr. David Yves Ghislain Delepine
Director
División de Ciencias e Ingenierías
Universidad de Guanajuato
Campus León

Por medio de la presente le informo que he revisado la tesis que presenta el C. **Jesús Fuentes Aguilar** para obtener el grado de Doctorado en Física. El trabajo de tesis del C. Fuentes Aguilar se titula: *On Non-Extensive Entropies: With Applications in Stochastic Dynamics and Information Theory*. En esta tesis doctoral el C. Fuentes Aguilar realiza algunas aplicaciones de una familia de entropías no-extensivas que no dependen de algún parámetro y dependen solo de la probabilidad. Se aplica esta variante de entropía a la compresión de datos en la teoría de la información y a algunos procesos biológicos que involucran procesos de difusión y transporte.

En términos generales, el trabajo realizado es de gran calidad y ha generado artículos de investigación ya publicados en revistas indexadas. El estudiante maneja el material en forma sólida y fluida. La tesis está bien escrita y su estructura es clara y se lee fácilmente. No tengo ninguna duda de que la calidad del trabajo del Sr. Fuentes Aguilar cumple con los requisitos de manera sobresaliente para que se proceda con los tramites para que se realice la defensa de tesis.

Aprovecho la ocasión para enviarle saludos cordiales.



Dr. Héctor Hugo García Compeán
Departamento de Física
Cinvestav-IPN