The Shannon–McMillan Theorem Proves Convergence to Equiprobability of Boltzmann's Microstates

Arnaldo Spalvieri

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy; arnaldo.spalvieri@polimi.it

Abstract: This paper shows that, for a large number of particles and for distinguishable and non-interacting identical particles, convergence to equiprobability of the W microstates of the famous Boltzmann–Planck entropy formula $S = k \log(W)$ is proved by the Shannon–McMillan theorem, a cornerstone of information theory. This result further strengthens the link between information theory and statistical mechanics.

Keywords: Boltzmann–Planck entropy formula; Shannon–McMillan theorem, equiprobability of microstates

1. Introduction

Entropy plays a key role in communication and information theory and in thermodynamics. Starting from the concept of entropy, many authors in the past built bridges between information and communication theory and thermodynamics, such that all the textbooks of statistical mechanics and thermal physics have chapters or sections where the basics of information theory are introduced. A recent survey about the links between entropy in communications and in thermodynamics is [1]. Making a comprehensive review of the bibliography on this topic is out of the scope of the present paper. We limit ourselves to mentioning the pioneering work of Jaynes [2] and the attention that has been attracted in the past by the connections between information theoretic inequalities and the irreversibility of certain thermodynamical processes; see [3], Chapter 4 of [4,5].

The aim of this paper is to put the information-theoretic concept of *typical set* inside the framework of statistical mechanics, thus strengthening the link between the two disciplines. Typicality is not a new word in statistical mechanics. It has been introduced in [6] for systems that are not at the equilibrium to mean that these systems, in their spontaneous evolution, *typically* tend to the maximum entropy. The difference with [6] is that (non) equilibrium is not our concern and that the typicality we refer to is the one that is technically established in information theory by the Shannon–McMillan theorem. Specifically, this paper shows that the Shannon–McMillan theorem *proves* that the microstates of the famous Boltzmann–Planck entropy formula are equiprobable. This result is surprising because, since the times of Boltzmann, the entire scientific community has postulated the equiprobability of microstates. Today, all the most influential textbooks of statistical mechanics and thermal physics unanimously say that microstates' equiprobability is the central postulate of statistical mechanics; see for instance 4.2 of [7], chapter 6 of [8], 16.1 of [9], and 1.1 of [10]. In 3.4 of [11], the author not only postulates equiprobability but also expresses, in the passage that we hereafter quote, strong scepticism about the possibility of proving it.

It is, unfortunately, impossible to prove with mathematical rigor that the principle of equal a priori probabilities applies to many-particle systems. Over the years, many people have attempted this proof, and all have failed. Not surprisingly, therefore, statistical mechanics was greeted with a great deal of scepticism when it was first proposed in the late 1800s. One of the its main proponents, Ludwig Boltzmann, became so discouraged by

all of the criticism that he eventually committed suicide. Nowadays, statistical mechanics is completely accepted into the cannon of physics—quite simply because it works.

The outline of the paper is as follows. In Section 2, we introduce our model of the thermodynamical system, where the only assumption that we make is that the system is made by a large number of non-interacting and identical particles. We say *our model* because it is slightly different from the commonly accepted one even if, in the thermodynamical limit of infinite number of particles, our model becomes compatible with the standard one. Section 3 shows that, with the further assumption of distinguishable particles, the set of microstates accessible to the thermodynamical system is the information-theoretic typical set of the Shannon–McMillan theorem. A crucial property of the typical set is that the probability distribution of its elements converges to uniformity; therefore, the identification between the typical set and the set of accessible microstates proves convergence to equiprobability of the accessible microstates. Finally, in Section 4 we recap the main points of the paper and draw conclusions.

2. Notation and System Model

Let uppercase calligraphic characters, e.g., $\mathcal{X}$, denote discrete random variables; let the set $\{\mathcal{X}\}$ be the support set of the random variable, e.g.,

$$\{\mathcal{X}\} = \{x_0, x_1, \cdots\};$$

and let $|\mathcal{X}|$ be the number of elements of the support set. The probability $Pr(\mathcal{X} = x)$ of the event $\mathcal{X} = x$ is denoted with the shorthand $p(x)$,

$$Pr(\mathcal{X} = x) = p(x), \ \ \forall\, x \in \{\mathcal{X}\}.$$

The probability distribution of $\mathcal{X}$ is the deterministic function $p(\cdot)$ of the random $\mathcal{X}$ with

$$\{p(\mathcal{X})\} = \{p(x_0), p(x_1), \cdots\}$$

and

$$Pr(p(\mathcal{X}) = p(x)) = p(x), \ \ \forall\ x \in \{\mathcal{X}\}.$$

Consider a thermodynamical system made by a fixed number N of non-interacting identical particles with one degree of freedom each. The thermodynamical system is modeled here by the random vector

$$\mathcal{E}_1^N = (\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_N)$$

of the energies of the N particles. We assume that energy is quantized and that the random entries of the random vector are independent (non-interacting), identically distributed (identical) (i.i.d.) random variables (particles); that is,

$$p(\mathcal{E}_1^N) = \prod_{i=1}^{N} p(\mathcal{E}_i), \tag{1}$$

with

$$p(\mathcal{E}_1) = p(\mathcal{E}_2) = \cdots = p(\mathcal{E}_N).$$

Since particles are i.i.d., one of them represents all. The random energy of the generic particle, which we call $\mathcal{E}$, takes its values in the set $\{\mathcal{E}\}$ of the allowed energy levels

$$\{\mathcal{E}\} = \{\epsilon_0, \epsilon_1, \cdots\}. \tag{2}$$

The expected energy of the particle is

$$E = \sum_{\epsilon \in \{\mathcal{E}\}} \epsilon \, p(\epsilon).$$

The total random energy of the system is

$$\sum_{i=1}^{N} \mathcal{E}_i.$$

By the i.i.d. assumption and by the law of large numbers, when $N \to \infty$, the system's random energy divided by the number of particles, that is, the random mean energy, tends to the deterministic expected energy:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}_i = E. \tag{3}$$

Convergence of the limit can be in various senses. In this paper, we will focus our attention on the weak law of large numbers (Khinchin law); hence, we consider convergence of the limit (3) *in probability*, meaning that, for every $\eta > 0$,

$$\lim_{N \to \infty} Pr\left(\left| \frac{1}{N} \sum_{i=1}^{N} \mathcal{E}_i - E \right| > \eta \right) = 0. \tag{4}$$

The limit in (4) is interpreted by saying that, for every η, if N is sufficiently large, then the total energy of the system lies in a narrow interval $\pm N\eta$ around NE with high probability.

A comment is in order about the approach pursued in this paper, which is based on the expected energy per particle, in comparison to the classical approach of thermal physics and statistical mechanics where the total system energy is considered. The assumption that the system's total energy lies in a narrow interval is the basis of the classical *microcanonical ensemble* approach; see, for instance, chapters 1 and 2 of [10]. In the microcanonical ensemble approach, the system is assumed to have a large and fixed number of particles and to be closed. Since N is large and fixed, there is basically no difference between the expected energy per particle and the total energy divided by the number of particles. When the number of particles is fixed but not as large, the classical approach in statistical mechanics and thermal physics is that of the *canonical ensemble*; see, for instance, chapter 3 of [10]. In the canonical ensemble approach, the thermodynamical system is at the thermal equilibrium at a temperature T with a heat bath and interacts with it to maintain the equilibrium condition. One of the consequences of the interaction between the system and the heat bath is that the total energy of the system fluctuates; hence, it becomes *substantially random*. Due to the random nature of total system's energy, in the canonical ensemble approach, it can no longer be used to characterize the system, which is characterized instead by the temperature of the heat bath. While the system's total energy becomes substantially random and must be abandoned in favor of the temperature of the heat bath, the expected energy, which is deterministic, can be related in a deterministic way to the temperature of the heat bath by the appropriate energy–temperature relation. For instance, the high-temperature approximation for wholly kinetic energy is

$$E = \frac{kT}{2},$$

where T is in Kelvin degrees and k is Boltzmann's constant. Hence, although in the following we will limit ourselves to $N \to \infty$, we see that the expected energy per particle can be used to characterize both the microcanonical and the canonical ensembles.

3. Equiprobability of Accessible Microstates

When we become aware of the result of a random experiment, in our case, the energy level of a particle, the *surprise* that we experience is the following deterministic function of the random result:

$$-\log(p(\mathcal{E})),$$

where $\log(x)$ indicates the logarithm of x and the base of the logarithm depends on the context. In information theory, the base is 2, while in physics, the base is Euler's constant e. Being a function of a random variable, the surprise is itself a random variable. The expectation is Shannon's entropy, which we call H:

$$H = - \sum_{\epsilon \in \{\mathcal{E}\}} p(\epsilon) \log(p(\epsilon)). \tag{5}$$

If distinguishable particles, or indexed particles, are considered, an outcome of $\mathcal{E}_1^N$ is an *energy microstate*, or, simply, *microstate*. For instance, a microstate is

$$\mathcal{E}_1^N : \ \mathcal{E}_1 = \epsilon_5, \ \mathcal{E}_2 = \epsilon_{31}, \ \cdots, \ \mathcal{E}_N = \epsilon_{18}, \tag{6}$$

meaning that the energy of the first particle is ϵ_5, the energy of the second particle is ϵ_{31}, $\cdots$, and the energy of the N-th particle is ϵ_{18}. This definition of microstate matches the commonly accepted one; see, e.g., 1.1 of [10], where the microstates of the system are the independent solutions of the Schröedinger equation of the system whose eigenvalue is the total energy of the system. Due to the assumption of independency between particles, the system decouples into N independent equations whose individual solutions are the wave functions whose energy eigenvalues are the energy levels (2), the sum of the individual energy eigenvalues being equal to system's total energy. For now, we settle for the definition of microstate; we will return soon on the concept of accessible microstate and on the energy constraint that the system must comply with.

When we become aware of the microstate visited by the system, the surprise that we experience is

$$-\log(p(\mathcal{E}_1^N)).$$

The expectation of the above surprise is the Shannon entropy of the system; hence the Shannon entropy of the random vector $\mathcal{E}_1^N$:

$$- \sum_{\epsilon_1^N \in \{\mathcal{E}_1^N\}} p(\epsilon_1^N) \log(p(\epsilon_1^N)). \tag{7}$$

In statistical mechanics, (7) is called *Gibbs entropy*. By the i.i.d. assumption (1), the entropy (7) is equal to n times the Shannon entropy of one particle:

$$- \sum_{\epsilon_1^N \in \{\mathcal{E}_1^N\}} p(\epsilon_1^N) \log(p(\epsilon_1^N)) = NH;$$

see 2.6.6 of [4]. One immediately recognizes that

$$\lim_{N \to \infty} -\frac{1}{N} \log(p(\mathcal{E}_1^N)) = \lim_{N \to \infty} -\frac{1}{N} \sum_{i=1}^{N} \log(p(\mathcal{E}_i)) \tag{8}$$

$$= H, \tag{9}$$

where (8) is the assumption of independency between particles and (9) follows from the assumption of identically distributed random variables and from the law of large numbers. The equality between the leftmost term and the rightmost one in the above two equalities shows that, as $N \to \infty$, the distribution of the microstates that the system can visit converges to the uniform distribution, because H is a deterministic and fixed quantity,

independent of the specific microstate $\mathcal{E}_1^N$. The convergence of the limit (8) to (9) is true in various senses. Convergence in probability, that is

$$\lim_{N \to \infty} Pr\left(\left|\frac{1}{N}\log(p(\mathcal{E}_1^N))) + H\right| > \eta\right) = 0, \tag{10}$$

is referred to as the Asymptotic Equipartition Property (AEP) in chapter 3 of [4], while many authors refer to (10) as to the Shannon–McMillan theorem (here it is the probability that is equally partitioned, not the energy; hence this AEP has nothing to do with the classical energy equipartition property). Almost everywhere convergence, that strengthens convergence in probability, is known as the Shannon–McMillan–Breiman theorem [12]. Exactly as it happens with total system's energy, here, for sufficiently large but finite N, the surprise about the system's microstate lies in a narrow range $\pm N\eta$ around NH with high probability and, exactly as it happens with the expected energy per particle E, H, which is *per particle*, can be used to characterize the entropy both in the microcanonical and the in canonical ensemble approaches even if, in what follows, we will limit ourselves to $N \to \infty$.

3.1. Typical Set and Accessible Microstates

Starting from (10), we arrive at the definition of *typical set*, which is the subset $\{\mathcal{T}_1^N(\eta)\}$ of $\{\mathcal{E}_1^N\}$ made by the vectors $\mathcal{E}_1^N$ whose probability, for any η and for sufficiently large N, is in the narrow range

$$e^{-N(H+\eta)} \leq p(\mathcal{E}_1^N) \leq e^{-N(H-\eta)}. \tag{11}$$

The properties of the typical set are that the number of its elements is in the narrow range

$$e^{N(H+\eta)} \geq |\mathcal{T}_1^N(\eta)| \geq (1 - \eta)e^{N(H-\eta)} \tag{12}$$

and that the probability that the system visits one element of the typical set is

$$Pr(\mathcal{E}_1^N) \in \{\mathcal{T}_1^N(\eta)\}) > 1 - \eta; \tag{13}$$

see again [4]. Formulas (11)–(13) show that with high probability, the system visits, or, in the language of statistical mechanics, *makes access to*, one among $|\mathcal{T}_1^N(\eta)|$ microstates. In addition, the distribution of microstates converges to the uniform one as $N \to \infty$, because $|\mathcal{T}_1^N(\eta)|$ is basically the inverse of $p(\mathcal{E}_1^N)$, which is constrained to lie in the narrow range (11). Note that convergence to equiprobability of accessible microstates is a *consequence* of the i.i.d. assumption and of $N \to \infty$, hence, in the end, of the law of large numbers, not a *postulate*.

The AEP divides the support set of energy microstates into two subsets: the typical set, which is the set of accessible microstates, and its complement, which is the set of microstates that the system cannot access. The division can be appreciated by observing that

$$|\mathcal{E}| - e^H \geq 0,$$

with equality when the energy levels that the single particle can visit are equiprobable, hence

$$|\mathcal{E}_1^N| = |\mathcal{E}|^N \geq e^{NH}. \tag{14}$$

Writing, with some abuse of mathematics, (12) as

$$|\mathcal{T}_1^N(\eta)| \approx e^{NH},$$

we see that, when inequality (14) fits with equality, the size of the typical set is the entire support set made of $|\mathcal{E}|^N$ elements, while, when $|\mathcal{E}| > e^H$, the number of elements of the typical set is lower than $|\mathcal{E}|^N$, leading to the division of the support set into the two subsets mentioned above. In this case, the system will visit with high probability the typical set and will visit with vanishingly small probability the complement of the typical set to the support set.

We now return to accessibility and the energy constraint to comment again on the key difference between our approach and the standard one. In the standard approach, for a microstate to be accessible, the energy eigenvalue of the wave function that solves the system's Schröedinger equation of the system must be equal to the constraint of total energy that is imposed on the system, while here it is the probability distribution $\{p(\mathcal{E})\}$ that must have a prescribed expected energy per particle. Compatibility between our approach and the standard one is guaranteed because, when $N \to \infty$, the energy eigenvalue of the system divided by the number of particles *is forced* by the law of large numbers to be equal to the expected energy of one particle; see (3).

3.2. Boltzmann–Planck Entropy

The equality between the Boltzmann–Planck entropy and N times the Shannon entropy of one particle is widely accepted in textbooks of statistical mechanics and thermal physics; see, e.g., Equation (18.14) of [9]. However, while in the standard approach, both the Boltzmann–Planck entropy and the Shannon entropy are purely deterministic quantities, in our approach, the Boltzmann–Planck entropy is the random quantity that we are will introduce in (18), that, divided by the number of particles, becomes equal to the deterministic Shannon entropy when the number of particles tends to infinity. We hereafter discuss the key passage, the passage where our model becomes compatible with the standard combinatorial approach to the evaluation of the number of accessible microstates.

Let $\mathcal{N}(\epsilon)$ be the random number of particles that visit the generic energy level ϵ, with

$$\sum_{\epsilon \in \{\mathcal{E}\}} \mathcal{N}(\epsilon) = N. \tag{15}$$

Compatibility of the accessible microstates with constraints imposed on the system will be guaranteed by imposing compatibility of the set

$$\{\mathcal{N}(\epsilon_0), \mathcal{N}(\epsilon_1), \cdots\} \tag{16}$$

with the distribution $\{p(\mathcal{E})\}$. The random number of microstates is the multinomial coefficient

$$\mathcal{W} = \frac{N!}{\prod_{\epsilon \in \{\mathcal{E}\}} \mathcal{N}(\epsilon)!} \tag{17}$$

and the random Boltzmann–Planck entropy is

$$\mathcal{S} = k \log(\mathcal{W}) \tag{18}$$

Randomness of the set (16) and, as a consequence, of the number of microstates (17) and of the Boltzmann–Planck entropy (18) marks the difference between our approach and the standard, deterministic one. Exactly as in the case of the energy, once again the law of large numbers bridges the gap between randomness and determinism, thus making our *random* approach compatible with the *deterministic* standard approach, hence the key passage is

$$\lim_{N \to \infty} \frac{1}{N} \log(\mathcal{W}) = H. \tag{19}$$

Specifically, step by step, one has

$$\lim_{N\to\infty} \frac{1}{kN}\mathcal{S} = \lim_{N\to\infty} \frac{1}{N}\log(\mathcal{W})$$

$$= \lim_{N\to\infty} \frac{1}{N}\left(\log(N!) - \sum_{\epsilon\in\{\mathcal{E}\}}\log(\mathcal{N}(\epsilon)!)\right)$$

$$= \lim_{N\to\infty} \frac{1}{N}\left(N\log\left(\frac{N}{e}\right) - \sum_{\epsilon\in\{\mathcal{E}\}}\mathcal{N}(\epsilon)\log\left(\frac{\mathcal{N}(\epsilon)}{e}\right)\right) \tag{20}$$

$$= \lim_{N\to\infty} \frac{1}{N}\left(N\log(N) - \sum_{\epsilon\in\{\mathcal{E}\}}\mathcal{N}(\epsilon)\log(\mathcal{N}(\epsilon))\right) \tag{21}$$

$$= \lim_{N\to\infty} \sum_{\epsilon\in\{\mathcal{E}\}}\frac{\mathcal{N}(\epsilon)}{N}(\log(N) - \log(\mathcal{N}(\epsilon))) \tag{22}$$

$$= \lim_{N\to\infty} - \sum_{\epsilon\in\{\mathcal{E}\}}\frac{\mathcal{N}(\epsilon)}{N}\log\left(\frac{\mathcal{N}(\epsilon)}{N}\right)$$

$$= - \sum_{\epsilon\in\{\mathcal{E}\}} p(\epsilon)\log(p(\epsilon)) \tag{23}$$

$$= H,$$

where in (21) and (22) we use (15), (23) is the law of large numbers and (20) is Stirling's formula that, in the big $\mathcal{O}$ notation, is

$$\log(N!) = N\log\left(\frac{N}{e}\right) + \mathcal{O}(\log(N)),$$

and we put in (20)

$$\lim_{N\to\infty} \frac{\mathcal{O}(\log(N))}{N} = \lim_{N\to\infty} \frac{\mathcal{O}(\log(\mathcal{N}(\epsilon)))}{N} = 0, \ \forall\, \epsilon \in \{\mathcal{E}\}.$$

The law of large numbers is also invoked in [13] in a passage that is basically the same as our (23) to show the equivalence of the phase space formulation of Gibbs and Boltzmann–Planck entropies.

From (12) and (19) we conclude that, for every small $\eta > 0$,

$$\lim_{N\to\infty} \frac{1}{N}\log(\mathcal{W}) = \lim_{N\to\infty} \frac{1}{N}\log(|\mathcal{T}_1^N(\eta)|). \tag{24}$$

Therefore, not only is our definition of energy microstates exemplified in (6) the commonly accepted one, but also the logarithm of the number of accessible microstates that we obtain by our approach is, for $N \to \infty$, equal to the logarithm of the standard combinatorial definition (17) of accessible microstates.

4. Conclusions

This paper has demonstrated that there is no need to *assume* or *postulate* the equiprobability of accessible microstates, because convergence of microstates' distribution to uniformity is a consequence of the large number of particles, of the i.i.d. assumption and of the assumption of distinguishable particles. Specifically, Equations (11)–(13) show that convergence to equiprobability is a consequence of the AEP, and Equation (24) shows that the microstates we are dealing with are the microstates that are counted by the standard combinatorial approach.

It is worth observing that equiprobability has nothing to do with the thermal equilibrium and hence with entropy maximization. Actually, the only conditions for convergence to equiprobability are the assumptions that we made, which do not include thermal equi-

librium and/or entropy maximization; therefore, the distribution converges to the uniform one also if the system is not at the thermal equilibrium. What can change between a system at the equilibrium and a system that is not at the equilibrium when both are subject to the same constraints is the probability distribution of energy levels and, with it, Shannon's entropy and, with it again, the number of accessible microstates, which, however, tend to equiprobability. Since the accessible microstates also remain virtually equiprobable in systems that are not at the equilibrium, the entropy of these systems is still given by the Boltzmann–Planck entropy formula, provided that the number of accessible microstates is conveniently expressed through the Shannon entropy of the specific (non-equilibrium) probability distribution of the energy of the individual particle by (5), (12).

When N is not large, e.g., $N = 1$, microstates are in general not equiprobable, independently of the entropy of the system.

To summarize, the main result of this paper is the following:

- For large N and i.i.d. particles, the accessible microstates are always virtually equiprobable, and also for systems that are not at the thermal equilibrium;
- For small N, the accessible microstates are never equiprobable, and also for systems that are at the thermal equilibrium with a heat bath.

The important case of indistinguishable particles is not treated in the paper and is left to future research.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Martinelli, M. Photons, Bits and Entropy: From Planck to Shannon at the Roots of the Information Age. *Entropy* **2017**, *19*, 341. [CrossRef]
2. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *4*, 620–630. [CrossRef]
3. Merhav, N. Physics of the Shannon Limits. *IEEE Trans. Inform. Theory* **2010**, *9*, 4274–4285. [CrossRef]
4. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley Series in Telecommunications and Signal Processing; John Wiley and Sons: Hoboken, NJ, USA, 2006.
5. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **1961**, *5*, 183–191. [CrossRef]
6. Goldstein, S.; Lebowitz, J.L. On the (Boltzmann) entropy of non-equilibrium systems. *Phys. D Nonlinear Phenom.* **2004**, *193*, 53–66. [CrossRef]
7. Kardar, M. *Statistical Physics of Particles*; Cambridge University Press: Cambridge, UK, 2007.
8. Huang, K. *Statistical Mechanics*; Wiley: Hoboken, NJ, USA, 1987.
9. Sekerka, R.F. *Thermal Physics: Thermodynamics and Statistical Mechanics for Scientists and Engineers*; Elsevier: Amsterdam, The Netherlands, 2015.
10. Pathria, R.K.; Beale, P.D. *Statistical Mechanics*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2011.
11. Fitzpatrick, R. *Thermodynamics and Statistical Mechanics*; World Scientific: Singapore, 2020.
12. Breiman, L. The individual ergodic theorem of information theory. *Ann. Math. Stat.* **1957**, *28*, 809–811. [CrossRef]
13. Frigg, R.; Werndl, C. A guide for the perplexed. *Probab. Phys.* **2011**, *2*, 115.