An Informational Proof of H-Theorem

Vincenzo Manca

Department of Computer Science and Center for BioMedical Computing, University of Verona, Verona, Italy

Email: vincenzo.manca@univr.it

Abstract

After a historical reconstruction of the main Boltzmann's ideas on mechanical statistics, a discrete version of Boltzmann's H-theorem is proved, by using basic concepts of information theory. Namely, H-theorem follows from the central limit theorem, acting inside a closed physical system, and from the maximum entropy law for normal probability distributions, which is a consequence of Kullback-Leibler entropic divergence positivity. Finally, the relevance of discreteness and probability, for a deep comprehension of the relationship between physical and informational entropy, is analyzed and discussed in the light of new perspectives emerging in computational genomics.

Subject Areas

Modern Physics

Keywords

Thermodynamic Entropy, H-Theorem, Information Entropy, Entropic Divergence

1. Introduction

Entropy appeared in physics, in the context of thermodynamics. However, after Boltzmann, it became a crucial concept for the statistical mechanics and for the whole physics. Since Shannon's paper of 1948 [1], information entropy was the basis of information theory on which current IT technology relies. It is really impressive that physical and informational entropies are essentially the same thing, and probably this common essence is so far not completely and properly understood. Physicist John Archibald Wheeler claimed in his speculations [2]: "It from bit" (all things physical are information-theoretic in origin). Namely, many open problems raised by modern physics will probably be clarified in a more general framework where informational concepts will be the main part of new physical theories. Entropy is directly related to information, probability,

complexity, and time. Boltzmann's investigations were aimed at reducing thermodynamical entropy to mechanical concepts, by integrating Newton's mechanics with statistical distributions. The focus of this reduction was the determination of a microscopical representation of entropy through which explaining the arrow of time implied by thermodynamics. This explanation was the so called H-Theorem that Boltzmann introduced in 1872 [3], showing that physical time is oriented as the decrease of the H-function (the entropy, apart from additive and multiplicative constants and the change of sign). A satisfactory proof of this property was never completely found by Boltzmann.

In this paper, we prove a discrete version of the H-Theorem by using basic physical and informational concepts. Namely, Boltzmann's difficulties to prove the H-Theorem are due to the information-theoretic nature of this theorem, more than fifty years before the foundation of Information Theory.

The present paper is mainly of methodological nature. In this perspective any discipline based on information-theoretic concepts may benefit from the analysis developed in the paper as it concerns with the link between the physical and the informational perspectives of entropy.

2. From Physical to Informational Entropy

In 1824, Sadi Carnot wrote a book about the power of fire in producing energy [4]. It was known that heat can provide mechanical energy. Moreover, as it was experimentally later shown by Joule [3], mechanical energy can be entirely transformed into heat (first law of thermodynamics). However, it is impossible entirely transforming heat into mechanical energy. In fact, this is a way to formulate the second law of thermodynamics, which is also equivalent to the impossibility of spontaneous processes where heat passes from a colder body to a hotter body. Carnot introduced some ideal heat engines, called reversible engines, reaching the optimal efficiency in heat-work transformation. In this way, he proved a theorem giving an evaluation of the heat quantity that necessarily cannot be transformed into mechanical energy by heat engines. Namely, when an engine M takes a quantity Q_2 of heat from a heat source (boiler) at constant temperature T_2, then a quantity Q_1 has to be released to a colder body (condenser) at temperature $T_1 (T_2 > T_1)$, and only the difference $Q_2 - Q_1$ can be transformed into mechanical work. For reversible engines working between T_2, T_1, the released heat quantity Q_1 reaches the minimum (positive) value, and the following equation holds [5]:

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2} \tag{1}$$

therefore, if we denote by S the heat quantity Q_1 when T_1 is the unitary temperature and $T_2 = T$ we obtain:

$$S = Q/T . \tag{2}$$

S corresponds to a thermodynamical quantity, later called by Clausius [3] the entropy (of M), corresponding to the minimum heat quantity that necessarily a

heat engine, working between temperatures T and 1 has to release (to the unitary condenser) and that cannot be transformed into mechanical work.

The famous formulation of the second law of thermodynamics, by means of entropy S, asserts that in a closed system (that does not exchange energy with the external environment) the entropy cannot decrease in time.

In this scenario, in 1870s years, Ludwig Boltzmann started a research aimed at explaining the second law of thermodynamics in terms of Newtonian mechanics [6] [7] [8]. The main question was: "Where does time arrow came from?". In fact, in mechanics all the laws are symmetric with respect to time and the same equations tell us what happens in the future, but also what happened in the past. In no equation there is an explicit indication about the direction of events in time (the first Chapter of Wiener "Cybernetics" [9] is devoted to the different notions of time in Newtonian mechanics and in biology).

The first step of Boltzmann's project was a mechanical formulation of entropy. This formulation can be found by starting from the fundamental law of ideal gases, where P is the pressure, V the volume, T the absolute (Kelvin) temperature, N the number of gas moles, and R is the gas constant:

$$PV = NRT. \qquad (3)$$

If we pass from the gas moles N to the number of molecules n in the gas (by the relation $Na = n$ where a is the Avogadro constant), we get an equivalent formulation, where $k = Ra$ is now called the Boltzmann constant:

$$PV = nkT. \qquad (4)$$

Now, let us assume that the gas takes some heat by expanding from a volume V_1 to a volume V_2. Then, the quantity Q of this heat is given by:

$$Q = \int_{V_1}^{V_2} P \mathrm{d}v \qquad (5)$$

and by expressing P according to Equation (3), we get:

$$Q = \int_{V_1}^{V_2} nkT/V \mathrm{d}v = nkT \int_{V_1}^{V_2} 1/V \mathrm{d}v = nkT \left(\ln V_2 - \ln V_1 \right). \qquad (6)$$

Let assume to start from a unitary volume $V_0 = 1$. If in Equation (6) $V_1 = V_0$, $V = V_2$ and T is moved to the left member, then we obtain:

$$Q/T = nk \ln V \qquad (7)$$

that, according to Carnot's equation (2), gives:

$$S = nk \ln V \qquad (8)$$

that is:

$$S = k \ln V^n \qquad (9)$$

where V^n expresses the number of possible ways of allocating the n molecules in V volume cells. The passage from constant R to constant k and from N moles to n molecules, accordingly, is crucial to the microscopic reading of the formula (very often this is not adequately stressed when Boltzmann's argument is analyzed).

We can assume that the gas is spatially homogeneous, that is, the same number of molecules are in any volume cell, so that spatial positions of cells do not matter. Therefore, an allocation of n molecules in V volume cells, is completely determined by the (molecule) velocities allocated in each volume cell, apart from multiplicative constants (factorials expressing the indiscernibility of molecules and cells), which in logarithmic terms are additive constants that are omitted. In conclusion, velocities partition the n molecules of gas in a number of different velocity classes: the intervals $v_1 \pm \Delta, v_2 \pm \Delta, \cdots, v_m \pm \Delta$ (with Δ a fixed increment value, and $n_1, n_2, \cdots, n_m$ the numbers of molecules having velocities in the $v_1, v_2, \cdots, v_m$ centered intervals, respectively). Hence, all the number of allocations of n molecules in the volume V corresponds to the number of different ways W that n molecules can be distributed into m different velocity classes (the number of different micro-states associated to a given thermo-dynamic macro-state). Whence, the final famous equation is obtained:

$$S = k \ln W. \tag{10}$$

Equation (10) is reported in Boltzmann's tomb in Wien. In this form, the equation was later given by Max Planck, who followed Boltzmann's approach in his famous conference on December 14, 1900, from which Quantum Theory emerged [10]. The importance of this microscopic formulation of Entropy is the statistical approach that stemmed from it and that became crucial in the whole physics of twentieth century.

The computation of W was obtained from Boltzmann by means of the so called Maxwell-Boltzmann statistics, related to the "Wahrscheinlichkeit" principle, and can be described by the following deduction. Namely, Let us consider all the molecules within the same velocity class as undistinguishable, then the different distributions of molecules into the m velocity classes are given by the following multinomial expression:

$$W = \frac{n!}{n_1! n_2! \cdots n_m!} \tag{11}$$

therefore, by Equation (10):

$$S = k \ln \frac{n!}{n_1! n_2! \cdots n_m!} \tag{12}$$

that is, by using Stirling approximation $n! \approx \sqrt{2\pi n}\, n^n / e^n$ [11], $\ln n! \approx n \ln n$ and we get:

$$S = kn \ln n - k \left(n_1 \ln n_1 + n_2 \ln n_2 + \cdots n_m \ln n_m \right)$$

if, for $i = 1, \cdots, m$, we express n_i by means of $p_i = n_i / n$, that is, $n_i = n p_i$, then we get (remember that $p_1 + p_2 + \cdots + p_m = 1$):

$$S = kn \ln n - k \left(n p_1 \ln n p_1 + n p_2 \ln n p_2 + \cdots + n p_m \ln n p_m \right)$$

$$S = kn \ln n - kn \left[p_1 \left(\ln n + \ln p_1 \right) + p_2 \left(\ln n + \ln p_2 \right) + \cdots + p_m \left(\ln n + \ln p_m \right) \right]$$

$$S = kn \ln n - kn \ln n \left[p_1 + p_2 - + \cdots + p_m \right] - kn \left(p_1 \ln p_1 + p_2 \ln p_2 + \cdots + p_m \ln p_m \right)$$

$$S = -kn\sum\nolimits_{i=1}^{m} p_i \ln p_i.$$

A discrete version of function H can be defined as:

$$H = \sum_{i=1}^{m} n_i \ln n_i \tag{13}$$

hence the equations above show that, apart from additive and multiplicative constants, $S = -H$, therefore the second law of thermodynamics, asserting the non-decrease of entropy (for closed systems) is equivalent to the law of non-increasing H function.

This was the first time that probability become an essential conceptual framework for physical representations. Surely, Maxwell's anticipations on gas velocity distribution and on the microscopic interpretation of temperature in terms of average velocity [10] were fundamental to Boltzmann's speculations, but in his approach the probabilistic setting is intended as an essential key for explaining what otherwise was impossible to explain. The forthcoming physics will be strongly based on probability and almost all the important achievements of quantum physics would have been impossible without an extensive and innovative use of probability [10].

If we consider the following informational entropy H_S introduced by Shannon [1] for a discrete probability distribution $p = \{p_i\}_{i=1,m}$ (its extension to continuous probability distributions is obtained by replacing sum with integral):

$$H_S = -\sum_{i=1}^{m} p_i \lg p_i \tag{14}$$

then, it results from equations above that H_S coincides with the thermody-namical entropy apart from a multiplicative constant.

The next Boltzmann's result was the "H-Theorem", which he tried to prove since 1872. However, all his attempts of finding a satisfactory proof of this theorem were not completely successful. The reason of these failures is the information-theoretic character of H-Theorem. Therefore, we need to enter in this theory that started in 1948 with the famous Shannon's paper [1] "A mathematical theory of communication".

3. Shannon's Entropy

Shannon's idea was entirely probabilistic. In fact, we can measure the information of an event e by means of a function $F(p_e)$ of its probability p_e, because the more rare an event is, the more we gain information when it occurs. Moreover, this function has to be additive for events that are independent, that is:

$$I(e_1, e_2) = I(e_1) + I(e_2) \tag{15}$$

if p_{e_1}, p_{e_2} are the respective probabilities of e_1, e_2, this means that the function $F(p_e) = I(e)$ has to verify:

$$F(p_{e_1} \times p_{e_2}) = F(p_{e_1}) + F(p_{e_2})$$

and the simplest functions satisfying this requirement are the logarithmic functions. Therefore, if we use the logarithmic in base 2, denoted by $\lg$, as it is customary in Information Theory, then we can define the information of an event e with probability p_e as:

$$I(e) = -\lg(p_e). \tag{16}$$

With this probabilistic notion of information, given a discrete probability distribution $p = \{p_i\}_{i=1,k}$, also called by Shannon an *information source* (a set where a probabilities are associated to its elements), information entropy expresses the probabilistic mean of the information quantities associated to the events $\{e_i\}_{i=1,k}$, with respect to the probability distribution p.

An anecdote reports that when Shannon asked John von Neumann to suggest him a name for the quantity S, then von Neumann promptly answered: "Entropy. This is just entropy", by adding that with this name the success of information theory was quite sure, because only few men knew exactly what entropy was.

Another remarkable fact is the paradoxical nature of entropy, due to its intrinsic probabilistic nature. This paradox is apparent at the beginning of Shannon's paper [1]. In fact, Shannon says that he is searching for a measure of the Entropy (information) or uncertainty of events. But how can we consider these notions as equivalent? It would be something like searching for a measure of knowledge or ignorance. Can we reasonable identify a measures for opposite concepts? The answer to these questions can be found in the intrinsic orientation of events in time. When an event e can happen with a probability p, we can measure its information by a function of its *a priori* uncertainty p, but after it happens, we gain, *a posteriori*, the same amount of information associated to p, because *a priory* uncertainty was transformed into a gain of certainty. The same kind of opposite orientation is always present in all informational concepts and it is often a source of confusion when the perspective of considering them is not clearly defined [12].

4. Boltzmann's H-Theorem

A concept related to entropy is the entropic divergence between two probability distributions p, q:

$$D(p\|q) = \sum_{x \in X} p(x)\lg(p(x)/q(x)). \tag{17}$$

The following proposition (see [13] for a proof) is a basic property of entropic divergence.

Proposition 1. $D(p\|q) \geq 0$ *for any two probability distributions.*

In an ideal gas the collisions between molecules are elastic, that is, the sum of the kinetic energies of two colliding molecules does not change after collision.

Proposition 2. *In an isolated ideal gas the variance of velocity distributions is constant in time.*

Proof. If the gas is isolated then its temperature is constant, and as Maxwell

proved [3] its temperature is uniquely related to its average velocity. Therefore, the average kinetic energy is constant in time, and being the gas ideal, its whole kinetic energy is constant, that is, also the second order momentum of velocity distribution is constant. Therefore *a fortiori* the variance is constant too. □

Let $N(x)$ denote the normal probability distribution $N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$, and $S(f)$ be the Shannon entropy of a probability distribution f (here $D(f\|N)$ is the continuous version of Kullback-Leibler divergence). The following propositions are proven in [13], but we report their proofs for reader's convenience.

Proposition 3.

$$S(N) = \frac{1}{2}\ln(2\pi e\sigma^2) \tag{18}$$

Proof. ([13]).

$$S(N) = -\int_{-\infty}^{+\infty} N(x)\ln N(x)\,dx$$

$$= \int_{-\infty}^{+\infty} -N(x)\ln\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\,dx$$

$$= \int_{-\infty}^{+\infty} -N(x)\left[-\frac{x^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2}\right]dx$$

$$= \int_{-\infty}^{+\infty} \frac{N(x)x^2}{2\sigma^2}\,dx + \ln\sqrt{2\pi\sigma^2}\int_{-\infty}^{+\infty} N(x)\,dx$$

$$= \frac{E(x^2)}{2\sigma^2} + \ln\sqrt{2\pi\sigma^2} \times 1$$

$$= \frac{1}{2} + \frac{1}{2}\ln 2\pi\sigma^2$$

$$= \frac{1}{2}\ln 2\pi e\sigma^2.$$

Proposition 4. *Normal distributions are those for which entropy reaches the maximum value in the class of probability distributions having the same variance.*

Proof. ([13]). Let f denote any probability distribution of variance $\text{Var}(f) = \sigma^2$.

$$D(f\|N) = \int_{-\infty}^{+\infty} f(x)\ln\frac{f(x)}{N(x)}\,dx$$

$$D(f\|N) = \int_{-\infty}^{+\infty} f(x)\ln f(x)\,dx - \int_{-\infty}^{+\infty} f(x)\ln N(x)\,dx$$

$$D(f\|N) = \int_{-\infty}^{+\infty} f(x)\ln f(x)\,dx - \int_{-\infty}^{+\infty} f(x)\ln\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\,dx$$

$$D(f\|N) = -S(f) - \int_{-\infty}^{+\infty} f(x)\ln e^{-\frac{x^2}{2\sigma^2}}\,dx + \ln\sqrt{2\pi\sigma^2}\int_{-\infty}^{+\infty} f(x)\,dx$$

$$D(f\|N) = -S(f) + \frac{1}{2\sigma^2}\int_{-\infty}^{+\infty} f(x)x^2\mathrm{d}x + \frac{1}{2}\ln 2\pi\sigma^2 \times 1$$

$$D(f\|N) = -S(f) + \frac{1}{2\sigma^2}\mathrm{Var}(f) + \frac{1}{2}\ln(2\pi\sigma^2) \quad \left(\mathrm{Var}(f) \le \sigma^2\right)$$

$$D(f\|N) \le -S(f) + \frac{1}{2} + \frac{1}{2}\ln(2\pi\sigma^2)$$

$$D(f\|N) = -S(f) + \frac{1}{2}\left(\ln \mathrm{e} + \ln(2\pi\sigma^2)\right)$$

$$D(f\|N) \le -S(f) + \frac{1}{2}\ln(2\pi\mathrm{e}\sigma^2) \quad (D(f\|N) \ge 0 \text{ by Proposition 1})$$

therefore $S(f) \le S(N)$. $\square$

Proposition 5. *In an isolated ideal gas, for each cartesian component, the velocity distribution, when normalized as a probability distribution, tends, in time, to reach the normal distribution* (*of a given variance*).

Proof. The proposition is a consequence of the central limit theorem [11]. In fact, if molecules collide randomly, the velocity of each molecule is a random variable. Let us consider the velocity intervals $v_1 \pm \Delta, v_2 \pm \Delta, \cdots, v_m \pm \Delta$ (of radius Δ). Then, a boolean random vector $X_{g,t} = (x_1, x_2, \cdots, x_m)$ can be associated to each gas molecule g and to any time point t taking values in a set of discrete time steps, in such a way that, if at time t molecule g has velocity in the interval $v_i \pm \Delta$, then $x_i = 1$ and $x_j = 0$, for every $j \ne i$, $1 \le j \le m$. In this way, the velocity distribution is the the the sum $\sum_{g,t} X_{g,t}$. Therefore, a large number of random variables are cumulated in the cartesian components of velocity distribution. $\square$

From the previous propositions Boltzmann's *H*-theorem follows.

Proposition 6. (H-Theorem) *In an isolated ideal gas the H function cannot increase in time.*

Proof. By Proposition 5, velocities tend to distribute according to a normal law, and they keep the variance constant, according to Proposition 2. But, according to Proposition 4, this distribution is that one having the maximum entropy in the class of all probability distributions with a given variance. $\square$

4.1. Pythagorean Recombination Game

A simple experimental validation of the above proposition can be obtained with a simple game, which we call *Pythagorean Recombination Game* (PRG), based on the following rules [14]. Chose a "large" multiset G of random numbers (numbers can occur in G with repetitions) and execute, at each step, the following operations:

1) Randomly choose two numbers a, b in G;

2) Randomly split a into a_1 and a_2, in such a way that $a = \sqrt{a_1^2 + a_2^2}$;

3) Randomly split b into b_1 and b_2, in such a way that $b = \sqrt{b_1^2 + b_2^2}$;

4) Replace in G the pair a, b with the pair $a' = \sqrt{a_1^2 + b_2^2}$, $b' = \sqrt{b_1^2 + a_2^2}$.

The idea of PRG is that of representing a two-dimensional gas where any

number is a velocity and the rules 2) and 3) define a collision axis and the exchange the velocity component with respect to this axis. If we play this game for a sufficiently number of steps and we compute the *H*-function at each step, we can easily check that H approaches to a minimum value, and at same time, the distribution of numbers (velocities within some intervals) clearly approximates to a χ distribution of freedom degree 2. This distribution corresponds to that of a random variable $\sqrt{X^2+Y^2}$ where both X,Y follow a normal distribution.

Proposition 7. *In Pythagorean Recombination Game H-function does not increase and the distribution of velocity components tend to a normal distributions as the game goes on.*

Proof. All the hypotheses of Proposition 6 are verified in PRG. In fact, collisions are elastic, because after each "collision" the kinetic energy remain unchanged. Moreover, the system is isolated because no energy is exchanged with the external world, and the numbers of colliding velocities and steps can be assumed to have values for which "large number laws" are acting in the process of velocity recombinations during the game. $\square$

The proposition above was experimentally verified by a suitable Matlab program implementing PRG (for suitable numerical parameters). **Figure 1** shows the shape of velocity distributions after only 1000 steps. A detailed description of this experiment, with the data and the obtained curves can be found in [14].

4.2. Entropy Generalizations and Applications

The schema of *H*-function definition from Boltzmann's formula $S = k \lg W$ is well known studied and presented in a lot of papers of physical and informational entropy. But there is an hidden aspect that has never properly investigated. The H function that is proportional to H_S Shannon entropy coincides, up to Stirling approximation, with:

$$H_S = \lg \frac{n!}{\Pi\{\{n_i!\}_{1\le i\le m}}. \tag{19}$$
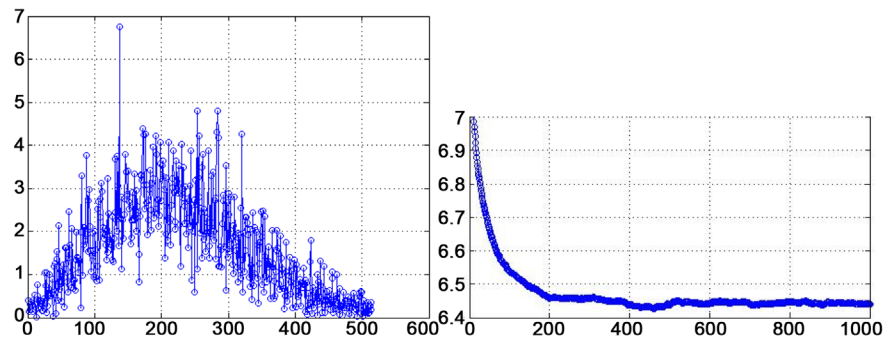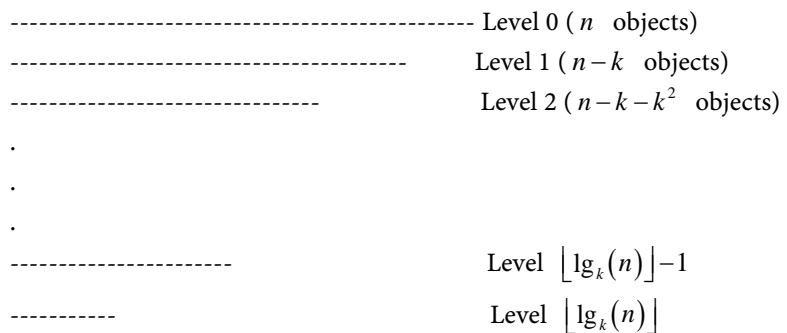

Figure 1. Left: Velocities after 1000 steps of PRG (50 collisions per step, 4000 particles, range of initial random velocities 500 - 600, level of approximation 2 ($v = v' \pm 2$). The shape of χ distribution of two freedom degrees appears. Right: H-function along the steps.

Here we briefly suggest a possible generalization of entropy that could be useful in the analysis of entropy for finite structures, that is something very relevant in the context of Machine Learning approach and computational genomics investigations [15] [16]. First, let us give a proposition about digital representability.

Proposition 8 (Digital Information Bounds) *The total number* $\text{digit}_k(n)$ *of digits necessary to encode* n *objects as strings, over an alphabet of* k *digits, verifies the following condition:*

$$n\left(\lfloor \lg_k(n) \rfloor + 1\right) \geq \text{digit}_k(n) \geq n\left(\lfloor \lg_k(n) \rfloor - 1\right). \tag{20}$$

Proof. We present a simple geometric proof of this proposition. Assume to place the n object we want to encode over k digits along a top line at level 0 (encoding n with a segment of n unitary lengths). Under it, we place shorter lines such that the line of level i has the length of the line above it minus k^i. We can depict this by means of the following line arrangement (here lengths are not proportional to the values):

```
------------------------------------------------ Level 0 ( n  objects)
-------------------------------------                Level 1 ( n − k  objects)
--------------------------------                     Level 2 ( n − k − k²  objects)
.
.
.
----------------------                               Level  ⌊lg_k(n)⌋ − 1
-----------                                          Level  ⌊lg_k(n)⌋
```

Now, the minimum number of digits that we need for encoding n objects is given by the sum of the lengths of all the lines of the arrangement. In fact, first line counts the initial number of digits that we need to assign at least one digit to each of the n objects. The second line assigns $n - k$ digits to the objects that need at least two digits to be encoded. Then, iteratively, at the $\lfloor \lg_k(n) \rfloor + 1$ step we add the last number of digits for the objects encoded with strings of maximum length. More formally we have that:

$$\text{digit}_k(n) = \sum_{i=0}^{\lfloor \lg_k(n) \rfloor} F_n(i) \tag{21}$$

where, for $0 \leq i \leq \lfloor lg_k(n) \rfloor$, $F_i(0) = n$, $F_i(i+1) = F(i) - k^{i+1}$.

Therefore, the exact number of digits is given by Equation (21), while the condition stated by the proposition follows from the fact that the number of digits used from level 0 to level $\lfloor \lg n \rfloor$ is surely less than $n\left(\lfloor lg_k(n) \rfloor + 1\right)$, while the number of digits used from level 0 to level $\lfloor \lg n \rfloor$ is surely greater than $n\lfloor \lg n \rfloor$ minus the missing parts (the gap with respect to n) of levels from 1 to $\lfloor \lg n \rfloor - 1$, and this missing part provides a sum less than n (as it easily results by evaluating $k + (k + k^2) + \cdots + (k + k^2 + \cdots + n^{\lfloor \lg n \rfloor - 1})$). Therefore, the number of digits used from level 0 to level $\lfloor \lg n \rfloor - 1$ is greater than $n\lfloor \lg n \rfloor - n$, so the condition is proved. □

The proof of *H*-Theorem given in the present paper substantially differs from classical proofs of the same type. In particular, Jaynes [17] developed a reinterpretation of Statistical Mechanics in terms of information-theoretic concepts. In this analysis key concepts are the *principle of maximum-entropy* in the probability estimation (MEP) and Shannon's Asymptotic Equipartition Principle (AEP). In [18] a proof of H-theorem is given where the informational approach to physical analysis provides (in a restricted case) important consequences about the relationship between Boltzmann and Gibbs approaches to the microscopic representation of thermodynamical states. However, a number of important differences can be individuated between papers [17] [18] and the analysis of the present paper. First, in both Jaynes' papers the theoretical setting is the classical continuous framework of physics. Moreover, the proof of H-theorem in [18] is not purely informational, but rather a physical proof where physical concepts are informationally reinterpreted, and the conclusions do not directly follow from properties of informational entropy. Other main differences are in the focus and in the perspectives, also due to the dates of papers [17] [18]. In fact, no efficient Montecarlo methods were available at that times, which could give simple numerical evidences of H-function behavior as coherent with Boltzmann intuition (exactly in the opposite direction of [18], where Gibbs' formulation is preferred). Contrarily, the Pythagorean Recombination Game given in Section 4.1 easily shows that the arrow toward the H-decrease is completely informational and corresponds exactly to Boltzmann's H-function trend. The discrete representation of H-function provides in a very simple way the essence of the theoretical proof and of its numerical validation. Although Boltzmann's intuition was surely based on a discretization of thermodynamical states, starting from Equation (9), the physical tradition prevailed in Boltzmann's search for physical arguments supporting the H-Theorem. Therefore, Boltzmann's revolution was in the crucial role of statistical distributions. Discreteness appeared as a theoretical necessity at the birth of Mechanical Statistics, but it becomes a key ingredient of physical modeling when Max Plank found the solution to the black body irradiation formula (starting from Boltzmann's $S = k \log W$ equation) [10], where the discrete nature of quanta appears linked to the energy probability distribution. In conclusion, both probability and discreteness have crucial roles in the comprehension of conceptual roots of statistical mechanics and were the conceptual seed of the subsequent development of the whole modern physics. This is the reason for which discrete proofs of H-Theorem provide a more general comprehension of the relationship between physical and informational entropy.

In many applications, where entropy is the empirical entropy computed from observed frequencies, the probability estimation is not the main aspect of the problem [15] [16]. For example, when we deal with genomic information as obtained by k-mer sequences of genomes [15] [19] other crucial aspects emerge such as: i) efficient methods for computing entropy, which is computationally prohibitive when k increases, and ii) the use of random genomes

for determining anti-random aspects that concern with the biological functions emerging during evolution. In conclusion, the discrete and computational perspectives are ingredients that are missing in classical discussions about information and physical entropies. But, exactly these perspectives are urgently requested in many contexts of application, where entropic concepts promise to disclose new possibilities of interpretation, especially in biological and medical fields. The next section wants just to give a broad idea of a possible discrete approach to entropic generalizations that, when combined with suitable algorithms, could provide useful tools for data analysis and interpretation [16].

In [15], a biological complexity measure, called biobit, is introduced for any genome $\mathbb{G}$, where two entropic components are combined: 1) the entropic component $EC(\mathbb{G})$ and 2) the anti-entropic component $AC(\mathbb{G})$. The component $EC(\mathbb{G})$ is the empirical entropy of genomes computed with respect to the word length $\lg_2(n)$ (n is the length of $\mathbb{G}$), whereas the anti-entropic component is the gap between $\lg_2(n)$ and $EC(\mathbb{G})$, where $\lg_2(n)$ results to be an upper bound to any empirical entropy of a genome long n (corresponding to a suitable Kullback-Leibler divergence). It results that along evolution the biobit measures of genomes increase. In a sense, this trend seems to contradict the law of increase for physical entropy. In fact, genomes are gigantic molecules that have to obey to the laws of physics, therefore, their biological complexity shows an evolutionary trend toward an opposite direction. The way of escaping from this paradox is given by the intrinsic symbolic nature of informational entropy of genomes, which transcends the physical limits of molecules. In fact, even if individual genomes are physical objects, the genomes of species are abstract symbolic sequences keeping their identity beyond their physical instances into individual genomes.

From the above Proposition it follows that the average length of strings encoding n objects is between $\lfloor \lg_k(n) \rfloor + 1$ and $\lfloor \lg_k(n) \rfloor - 1$, therefore $\log_k(n)$ can be assumed as (a good approximation to) the average amount of digital information that is intrinsically related to a set of n objects. Now, if we go back to Equation (19), in the light of proposition above, we discover that Shannon Entropy is proportional to the minimal digital information of the set of permutations of n objects into m classes that leave unchanged the cardinalities of these (disjoint) classes. This phenomenon can be easily generalized in many ways, by providing new simple methods for computing other kinds of entropies over finite structures.

Let us shorty outline a general notion of Structural Entropy directly related to Equation (19). Let us consider a finite structure Z (of some type) defined over a set A, called the support of Z. Let us consider the class $\mathbb{F}_A$ of 1-to-1 functions from A to A. The subet $\mathbb{F}_Z$ of $\mathbb{F}_A$ is the set of *iso-configurations* of Z sending Z in itself (or in an equivalent structure with respect to a suitable notion of structural equivalence). Therefore, we define the Structural Entropy of Z (of base k) as the average digital information of the isoconfigurations $\mathbb{F}_Z$ (see Proposition 8, $\|$ is set cardinality):

$$H_{\mathbb{F}}(Z) = \mathrm{digit}_k\left(|\mathbb{F}_Z|\right) / |\mathbb{F}_Z| \approx \lg_k\left(|\mathbb{F}_Z|\right). \tag{22}$$

The above definition tells us that Boltzmann's H-function corresponds to the Structural Entropy of a partition of n undistinguishable objects into m distinct sets, when $\mathbb{F}$ is the class all n-permutations. Topological entropy [20] is defined, following a similar idea, as the logarithm of a suitable set cardinality, but according to a limit process making hard its computational evaluation. New notions of structural entropies over finite structures, with efficient algorithms for computing them, are surely of great interest in all the cases where classical information entropy requires to be modified, for coping with the specific needs of many applicative contexts [21].

5. Conclusions

The arrow of time is a consequence of central limit theorem plus the informational law of maximum entropy of normal distributions (deduced from Kulback-Leibler entropic divergence positivity). For this reason, time irreversibility emerges when physical systems are complex enough and "large numbers" phenomena can act on them. This complexity, which is typical in living systems, is the origin of the intrinsic direction of biological time. Here a discrete version of H-theorem is proved, which shows the informational mechanisms on which it is based.

The proof of H-Theorem by means of concepts from probability and information theory has an interest going beyond the technical aspects of the proof. Namely, the focus is not on the result in itself (today many demonstrative approaches are known [22]), but it is in the discrete and informational approach here outlined, which is mainly aimed at motivating and suggesting investigations toward the interplay between information and energy in their specific mechanisms of interaction. This ambit remains mostly dense of enigmas, often related to concepts of quantum physics or of complex systems. In contexts where emergence phenomena occur, classical interpretations of the second laws seem to contradict the mechanisms of complexity, evolvability and autonomy related to new paradigms of informational inference and of distributed autonomous agent computations. The genomic laws given in [15] are an indication toward more comprehensive ways of considering entropic concepts, where the informational perspective seems to provide new keys for the analysis of complex phenomena.

Acknowledgements

References

[1] Shannon, C.E. (1948) A Mathematical Theory of Communication. *The Bell System Technical Journal*, **27**, 623-656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

[2] Wheeler, J.A. (1990) Information, Physics, Quantum: The Search for Links. In: Zu-

rek, W.H., Ed., *Complexity, Entropy, and the Physics of Information*, Addison-Wesley, Redwood City, 354-368.

[3] Brush, S.G. and Hall, N.S. (Eds.) (2003) The Kinetic Theory of Gases. An Anthology of Classical Papers with Historical Commentary. Imperial College Press, London.

[4] Carnot, S. (1890) Reflections on the Motive Power of Heat (English translation from French edition of 1824, with introduction by Lord Kelvin). Wiley & Sons, New York.

[5] Feynman, R.P. (1963) The Feynman Lectures on Physics. Vol. I, Part 2. Addison-Wesley Publishing Company, Inc., California Institute of Technology, Boston.

[6] Boltzmann, L. (1872) Weitere Studien uber das Wärmegleichgewicht unter Gasmolekulen, Sitzungsber. *Kais. Akad. Wiss. Wien Math. Naturwiss. Classe*, **66**, 275-370.

[7] Boltzmann, L. (1877) Über die Beziehung zwischen dem Zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. den Sätzen über das Wärmegleichgewicht, Sitzungsber. *Kais. Akad. Wiss. Wien Math. Naturwiss. Classe*, **76**, 373-435.

[8] Sharp, K. and Matschinsky, F. (2015) Translation of Ludwig Boltzmann's Paper "On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium". *Entropy*, **17**, 1971-2009. https://doi.org/10.3390/e17041971

[9] Wiener, N. (1948) Cybernetics or Control and Communication in the Animal and the Machine. Hermann, Paris.

[10] Planck, M. (1972) Planck's Original Papers in Quantum Physics. Annotated by H. Kangro, translated by Haar, ter D.; Brush, S. G. Taylor & Francis, London.

[11] Feller, W. (1968) An Introduction to Probability Theory and Its Applications. Wiley & Sons, New York.

[12] Brillouin, L. (1953) The Negentropy Principle of Information. *Journal of Applied Physics*, **24**, 1152-1163. https://doi.org/10.1063/1.1721463

[13] Cover, T.M. and Thomas, J.A. (1991) Elements of Information Theory. Wiley & Sons, New York. https://doi.org/10.1002/0471200611

[14] Manca, V. (2013) Infobiotics: Information in Biotic Systems. Springer, Berlin. https://doi.org/10.1007/978-3-642-36223-1

[15] Bonnici, V. and Manca, V. (2016) Informational Laws of Genome Structures. *Nature Scientific Report*, **6**, 28840. http://www.nature.com/articles/srep28840 https://doi.org/10.1038/srep28840

[16] Holzinger, A., Hörtenhuber, M., Mayer, C., Bachler, M., Wassertheurer, S., Pinho, A.J. and Koslicki, D. (2014) On Entropy-Based Data Mining. In: Holzinger, A. and Jurisica, I., Eds., *Knowledge Discovery and Data Mining*, Springer, Berlin, 209-226. https://doi.org/10.1007/978-3-662-43968-5_12

[17] Jaynes, E.T. (1957) Information Theory and Statistical Mechanics. *The Physical Review*, **106**, 620-630. https://doi.org/10.1103/PhysRev.106.620

[18] Jaynes, E.T. (1965) Gibbs vs Boltzmann Entropies. *American Journal of Physics*, **33**, 391-398. https://doi.org/10.1119/1.1971557

[19] Manca, V. (2016) Infogenomics: Genomes as Information Sources. In: Arabnia, H. and Tran, Q., Eds., *Emerging Trends in Applications and Infrastructures for Computational Biology*, Elsevier, Morgan, 317-324. https://doi.org/10.1016/b978-0-12-804203-8.00021-3

[20] Brin, N. and Stuck, G. (2002) Introduction to Dynamical Systems. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511755316

[21] Abhishek, S., Asok, R. and Shalabh, G. (2009) An Information-Theoretic Measure for Anomaly Detection in Complex Dynamical System. *Mechanical Systems and Signal Processing*, **23**, 358-371. https://doi.org/10.1016/j.ymssp.2008.04.007

[22] Cercignani, C. (1988) The Boltzmann Equation and Its Application. Springer, Berlin.

Open Access Library

Submit or recommend next manuscript to OALib Journal and we will provide best service for you:

- Publication frequency: Monthly
- 9 subject areas of science, technology and medicine
- Fair and rigorous peer-review system
- Fast publication process
- Article promotion in various social networking sites (LinkedIn, Facebook, Twitter, etc.)
- Maximum dissemination of your research work

Submit Your Paper Online: **Click Here to Submit**
Or Contact **service@oalib.com**