



# Law Without Law or “Just” Limit Theorems?

## Some Reflections About a Proposal of Wheeler’s

Sergio Caprara<sup>1,2</sup>  · Angelo Vulpiani<sup>1,3</sup>

Received: 13 February 2018 / Accepted: 16 August 2018 / Published online: 23 August 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

### Abstract

About 35 years ago Wheeler introduced the motto “law without law” to highlight the possibility that (at least a part of) Physics may be understood only following *regularity principles* and few *relevant facts*, rather than relying on a treatment in terms of fundamental theories. Such a proposal can be seen as part of a more general attempt (including the maximum entropy approach) summarized by the slogan “it from bit”, which privileges the information as the basic ingredient. Apparently it seems that it is possible to obtain, without the use of physical laws, some important results in an easy way, for instance, the probability distribution of the canonical ensemble. In this paper we will present a general discussion on those ideas of Wheeler’s that originated the motto “law without law”. In particular we will show how the claimed simplicity is only apparent and it is rather easy to produce wrong results. We will show that it is possible to obtain some of the results treated by Wheeler in the realm of the statistical mechanics, using precise assumptions and nontrivial results of probability theory, mainly concerning ergodicity and limit theorems.

**Keywords** Statistical mechanics · Probability · Typicality · Limit theorems

## 1 Introduction

In an influential paper [1] reporting his Oersted Medal Response at the joint APS-AAPT Meeting (25 January 1983) John Archibald Wheeler posed the following bold

---

✉ Sergio Caprara  
sergio.caprara@roma1.infn.it

<sup>1</sup> Dipartimento di Fisica - Università di Roma Sapienza, Piazzale Aldo Moro 5, 00185 Rome, Italy

<sup>2</sup> Institute for Complex Systems (ISC-CNR), UOS Sapienza, Piazzale A. Moro 5, 00185 Rome, Italy

<sup>3</sup> Centro Interdisciplinare Linceo “B. Segre”, Accademia dei Lincei, Via della Lungara, 10, 00165 Rome, Italy

problem: Is it possible to build part of Physics only following *regularity principles* (RP)? In other words, can we understand (at least a part of) Physics without referring to fundamental theories (e.g., classical or quantum mechanics)?

Wheeler proposed the “law without law” program. For instance in Ref. [1] he discussed three examples which, in his opinion, support the approach in terms of RP:

- (i) the Boltzmann–Gibbs distribution of the canonical ensemble;
- (ii) the universality of critical exponents;
- (iii) the statistical features of heavy nuclei.

Regarding point (i), Wheeler wonders: “How can stupid molecules ever be conceived to obey a law so simple and so general? [...] What regulating principle accomplished this miracle?”. And then he adds: “No small answer can ever hope to live up to a question so great”.

According to Wheeler [1], Physics may be divided into three Eras:

- Era number one (Copernicus, Kepler, and Galileo) is characterized by the discovery of the simplicity of motion;
- Era number two (starting with Newton) is marked by the concept of physical laws;
- Era number three (present time) is ruled by regulating principles, i.e., “chaos behind law”.

Here, the basic idea is that chaotic behavior and regulating principles can fruitfully cooperate and give rise to approximated laws. Let us note that Wheeler uses the term “chaos” in a wide sense, i.e., without the technical meaning of sensitivity to the initial conditions.

His message can be summarized in the motto

$$\left(\text{higgledy} - \text{piggledy}\right) + \left(\text{regulating principle}\right) = \left(\text{law of physics}\right). \quad (1)$$

Wheeler’s proposal can be seen as a part of a more general program whose famous slogan is “it from bit”, suggesting that the ultimate building block upon which we should base our knowledge of the World is information [2].

Deutsch [3], discussing Wheeler’s idea, puts forward the following dilemma: are the *regulating principles* analytic or synthetic propositions? He then tried to rewrite Wheeler’s proposal in the less inspiring form

$$\left(\text{stochastic laws}\right) \rightarrow \left(\text{approximate “deterministic laws”}\right). \quad (2)$$

Surely the proposal by Wheeler, even in the weaker Deutsch’ form, is quite vague, on the otherhand it is somehow part of a general approach to physics where the main ingredients are information and inference. Our main aim is a critical discussion of the limits of such an approach.

In particular, we will show that the examples discussed by Wheeler can appear as consequences of general regulatory principles only a posteriori. Actually they follow from results of probability theory that are far from trivial, basically the main ingredients are limit theorems, namely:

- The Boltzmann–Gibbs law follows from the microcanonical distribution, which is based on (some form of) the ergodic hypothesis.
- The universality of the statistical features of heavy nuclei is a sort of ergodicity: the single (large) nucleus is well described by its average properties.
- The universal properties of critical phenomena can be seen as a generalization of the central limit theorems.

In Sect. 2 we address the issue of the Boltzmann–Gibbs distribution, discussing an enlightening conjecture by Ulam, the notion of “bridge law”, the ergodic hypothesis and Khinchin’s approach, and the microcanonical distribution. In Sect. 3 we cast our reflections about Wheeler’s proposal into a wider perspective, discussing the case of universality in critical phenomena. In Sect. 4 we address the role and meaning of the maximum entropy principle, comment upon the actual role of probability in statistical physics. Some final remarks are found in Sect. 5.

## 2 The Boltzmann–Gibbs Distribution of the Canonical Ensemble

Let us briefly discuss some general aspects of the foundations of the statistical mechanics, in particular the role of ergodicity and of the many degrees of freedom in macroscopic objects [4–6].

We owe to Ludwig Boltzmann’s ingenuity two great ideas in modern Physics [5,7]:

- (A) the introduction of probability in the physical realm;
- (B) the bridge between the microscopic dynamics, and the macroscopic description of the physical world, which is based on a phenomenological theory, named thermodynamics.

Before discussing in some detail Boltzmann’s approach, we open a brief digression about an interesting conjecture by Ulam, which helps to clarify some aspects about statistical laws in physics, in particular the unavoidable role of the dynamics, or at least some aspects of the laws ruling the time evolution.

### 2.1 An Interlude About Stupid Molecules and Boltzmann’s Law

In his paper [1], Wheeler writes: *How can stupid molecules ever be conceived to obey a law so simple and so general? [...] What regulating principle accomplished this miracle?* And then he adds “No small answer can ever hope to live up to a question so great”. He discusses a model of  $N$  oscillators sharing quanta of energy, wondering how in the limit  $N \gg 1$  *stupid molecules, sharing energy higgledy-piggledy, nevertheless end on the average obeying Boltzmann’s law?* He then concludes that this is an example of “law without law” according to [1].

Let us briefly discuss an interesting conjecture of Ulam’s. Consider a (large) system with  $N$  particles and the following stochastic rule: at time  $t$  a pair  $(i, j)$  is selected at random, these two particles undergo a “collision” which preserves the total energy

but allows for a redistribution of the energy according to the rule:

$$E_i(t + 1) = X(t) [E_i(t) + E_j(t)], \quad E_j(t + 1) = [1 - X(t)] [E_i(t) + E_j(t)],$$

where  $X(t)$  is a random variable independent of  $X(t - 1)$ , and takes values in  $[0, 1]$ . Ulam conjectured that, if  $X$  is uniformly distributed, starting from any initial distribution of the energy, asymptotically Boltzmann’s probability density for the energy holds:  $P_E(E) = \alpha e^{-\alpha E}$  where  $\alpha = 1/E_0$  is determined by the mean energy  $E_0$  per particle at  $t = 0$ .

The above conjecture has been proved in a rigorous way by Blackwell and Mauldin [8]. In addition it is possible to generalize the result to the case of a generic probability distribution  $P_X(X)$ : independently of the initial distribution of the energy, asymptotically one has a convergence to an invariant probability density for the energy  $P_E(E)$  whose shape depends on  $P_X(X)$ . For instance, if  $P_X(X) = \delta(X - \frac{1}{2})$ , using the law of large numbers it is possible to show that  $P_E(E) = \delta(E - E_0)$  where  $E_0$  is the mean energy per particle at  $t = 0$ . In Fig. 1 we report numerical calculations obtained for a system with  $N = 50,000$  particles, undergoing 500,000 collisions. The last (asymptotic) distribution is smoothed by averaging over the last 500 realizations.

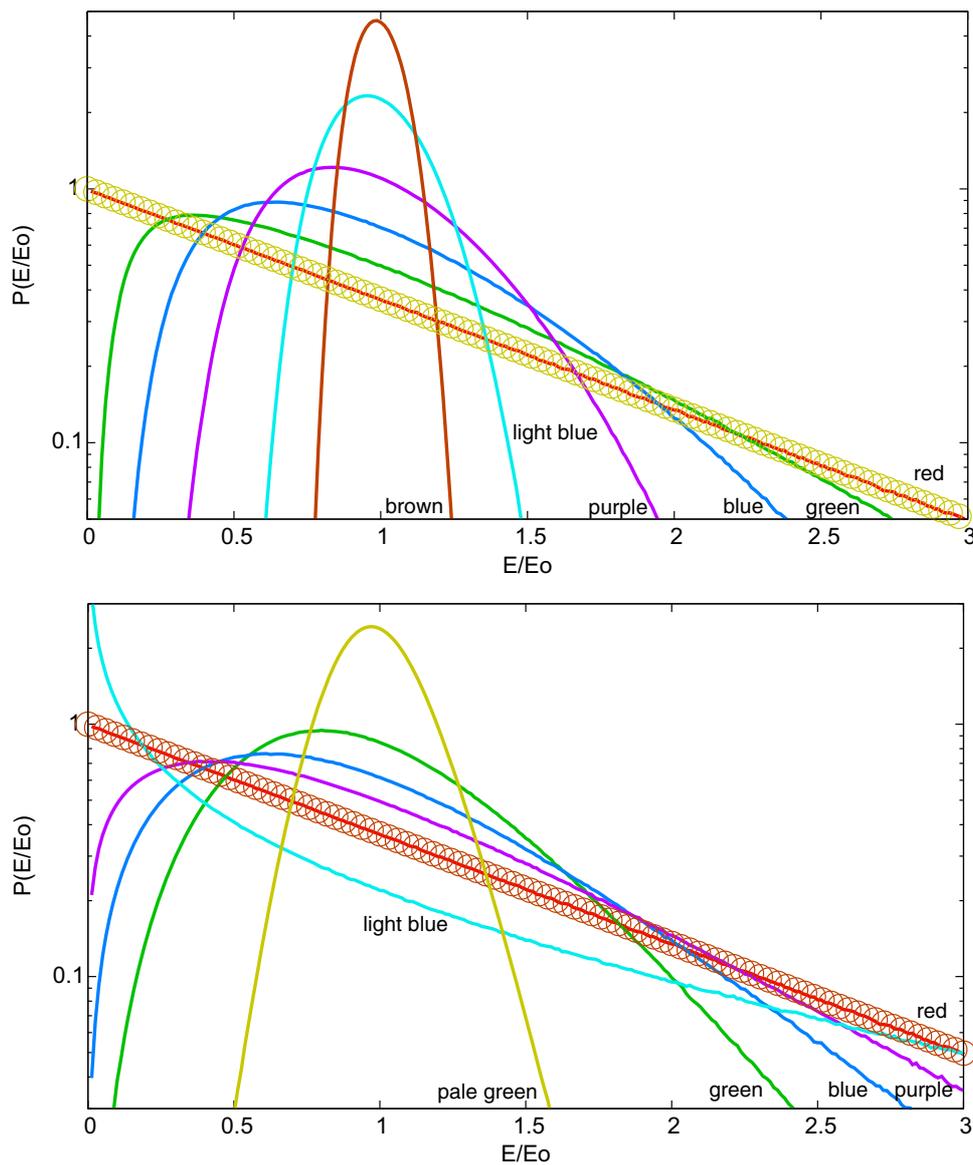
The results in Ref. [8], as well as the above simple numerical computations, show in an unambiguous way the relevance of the “correct rule”. Only for a specific  $P_X(X)$  one obtains the proper physical results: it is necessary to use a suitable stochastic law, while a generic “higgledy-piggledy rule” is not enough. The different results with varying  $P_X(X)$  are somehow related to Bertrand’s well-known paradox in probability theory [9]: the probability  $p$  that, considering an equilateral triangle inscribed in a circle, and picking at random a chord of the circle, the chosen chord is longer than a side of the triangle, depends on the way we interpret the prescription “at random”. For instance, if the two end points of the chord are uniformly distributed along the circumference, this probability is  $p = \frac{1}{3}$ ; if the midpoint of the chord is uniformly distributed along a randomly chosen radius,  $p = \frac{1}{2}$ ; if the midpoint of the chord is uniformly distributed over the entire circle,  $p = \frac{1}{4}$ . From these results one concludes that the term “at random” is ambiguous and therefore a precise prescription is needed to compute  $p$ .

We close this interlude noting that it is possible to reformulate Ulam’s approach in terms of velocity. A particularly interesting case is given by the granular gases where in the inelastic collision between two granular particles the momentum is preserved and in addition each particle interacts with a thermal bath. At variance with the elastic case the probability distribution of the velocity is not Gaussian and shows long tails [10] whose details depend on the parameters of the collision rules, e.g. the restitution coefficient and the characteristic time of the interaction particle-bath.

### 2.2 Bridge Law

The issue (B) can be summarized by the well-known relation

$$S = \kappa \log \mathcal{W}, \tag{3}$$



**Fig. 1** Top panel: Probability distribution of the energy  $P_E(E/E_0)$ , for various distributions densities  $P_X(X)$  which are constant over a given interval  $I \subset [0 : 1]$ , and zero outside it: red line,  $I = [0 : 1]$ ; green line,  $I = [0.1 : 0.9]$ ; blue line,  $I = [0.2 : 0.8]$ ; purple line,  $I = [0.3 : 0.7]$ ; light blue line,  $I = [0.4 : 0.6]$ ; brown line,  $I = [0.45 : 0.55]$ . Boltzmann’s probability density  $e^{-E/E_0}$  is highlighted by the empty circles. Bottom panel: Probability distribution of the energy  $P_E(E/E_0)$ , for various symmetric nonuniform distributions densities  $P_X(X) = P_X(1 - X)$ ,  $X \subset [0 : 1]$  (the expression for  $X \leq \frac{1}{2}$  is given): pale green line,  $P_X(X) = 10(2X)^9$ ; green line,  $P_X(X) = 3(2X)^2$ ; blue line,  $P_X(X) = 4X$ ; purple line,  $P_X(X) = \frac{3}{2}\sqrt{2X}$ ; light blue line,  $P_X(X) = 1/(2\sqrt{2X})$ ; the result for a uniform distribution is reported for comparison (red line). Again, Boltzmann’s probability density  $e^{-E/E_0}$  is highlighted by the empty circles (Color figure online)

which is engraved on Boltzmann’s tombstone in Vienna. Actually the above equation, usually called *Boltzmann’s law*, has been written by Planck [7].

Being  $H(\mathbf{Q}, \mathbf{P})$  the Hamiltonian of a given system whose particles are in a volume  $V$ ,  $S$  is the entropy (a thermodynamic quantity),

$$W(E, V, N) = \frac{1}{N!h^{3N}} \int_{E < H(\mathbf{Q}, \mathbf{P}) < E + \Delta E} d^{3N}\mathbf{Q} d^{3N}\mathbf{P} \tag{4}$$

is the “volume of phase space” (the space of all the coordinates  $\mathbf{Q}$  and the momenta  $\mathbf{P}$  specifying a microscopic configuration) accessible to a macroscopic configuration of  $N$  particles with energy in the interval  $[E, E + \Delta E]$ ,  $\kappa$  is a dimensional constant (the Boltzmann constant) and  $h$  is another dimensional constant (the Planck constant).

A simple, yet important, remark to the *bridge law* is in order here. Often, even in good textbooks (see, e.g., Ref. [11]), one can read that the bridge law between thermodynamics and mechanics is given by the equation

$$\langle \text{kinetic energy per particle} \rangle = \frac{3}{2}\kappa T,$$

relating the average kinetic energy of a particle with the absolute temperature  $T$ . This relation, which was already known to Daniel Bernoulli, holds in systems whose Hamiltonian contains the usual kinetic term  $\sum_k \mathbf{P}_k^2/2m$ . On the other hand it cannot be valid for a system with a generic Hamiltonian. As an important example we can mention those systems which can have negative temperatures, e.g., point-vortex systems in two-dimensional inviscid fluids [12].

Equation (3), with  $\mathcal{W}(E, V, N)$  given by Eq. (4), is the true *bridge law*, joining mechanics, i.e.  $\mathcal{W}(E, V, N)$ , and thermodynamics, i.e.  $\mathcal{S}(E, V, N)$ . Once  $\mathcal{W}(E, V, N)$  is known, we have the entropy and therefore the temperature  $T$  and the pressure  $P$  can be obtained through

$$\frac{1}{T} = \left( \frac{\partial \mathcal{S}}{\partial E} \right)_{V,N}; \quad \frac{P}{T} = \left( \frac{\partial \mathcal{S}}{\partial V} \right)_{E,N}.$$

### 2.3 The Ergodic Hypothesis

The issue (A) is rather subtle, and is still an open question. To shed light on this problem, let us consider a macroscopic system with  $N$  interacting particles, whose microscopic state is described by the vector  $\mathbf{X} \equiv (\mathbf{Q}_1, \dots, \mathbf{Q}_N; \mathbf{P}_1, \dots, \mathbf{P}_N) \in \mathbb{R}^{6N}$ . When an instrument measures a physical observable quantity, e.g., the pressure, it performs a time average of some function of  $\mathbf{X}$ ,

$$\frac{1}{T} \int_0^T \mathcal{A}(\mathbf{X}(t)) dt, \tag{5}$$

over the observation time  $T$ . Of course, in order to determine the time evolution  $\mathbf{X}(t)$ , the knowledge of the initial condition  $\mathbf{X}(t = 0)$  is needed. However, even in this case, the program can be accomplished only for systems composed of a reasonably small number of particles and is definitely hopeless when  $N$  grows as large as the Avogadro number (the number of  $\text{H}_2$  molecules contained in just 2 grams of molecular hydrogen).

Boltzmann’s ingenious idea consists in replacing the time average in Eq. (5) with a suitable average over phase space, assuming that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathcal{A}(\mathbf{X}(t)) dt = \int \mathcal{A}(\mathbf{X}) \rho_{micro}(\mathbf{X}) d\mathbf{X}, \tag{6}$$

where  $\rho_{micro}(\mathbf{X})$  is the microcanonical probability density. The limit  $\mathcal{T} \rightarrow \infty$  is necessary from a mathematical point of view [4,5] and it is physically well justified. The reason is that the usual observation time is much larger than the typical molecular times, e.g., the mean collision time in gases and liquids, or the oscillation periods in solids, which are  $O(10^{-10} \text{ s})$ . Once Eq. (6) is assumed to be valid in the limit of large  $N$ , in the presence of short-range interactions, it is straightforward to derive, for a system exchanging energy with a much larger external environment, the Boltzmann–Gibbs distribution for the canonical ensemble

$$\rho_{can}(\mathbf{X}) = \text{Const. } e^{-\beta H(\mathbf{X})},$$

with  $\beta \equiv 1/(\kappa T)$ .

Now, after the Fermi–Pasta–Ulam simulations [6,12] and the Kolmogorov–Arnol’d–Moser theorem [13], it is quite clear that, strictly speaking, the EH cannot be valid. On the other hand, in one of the most common and powerful method used in the numerical computation, the molecular dynamics [14], the time average of observables are computed assuming the validity of (6).

Why, in spite of the failure of the ergodicity, the results obtained with the Boltzmann–Gibbs distribution are valid? In particular why is it in agreement with the results of molecular dynamics? Let us briefly discuss the physical reasons which allow for the practical success of the standard computational techniques used in statistical mechanics.

Following Khinchin’s approach [15], it can be show that the EH is essentially true if (i) the system under study consists of a large number of particles, (ii) only “suitable” observable quantities are considered, and (iii) we allow for Eq. (6) to fail in a “sufficiently small region” (i.e., with a small probability). Khinchin was indeed able to show that for *sum functions* like

$$f(\mathbf{X}) = \sum_{k=1}^N f_k(\mathbf{Q}_k, \mathbf{P}_k) \quad (7)$$

the inequality

$$\text{Probability} \left( |\delta f(\mathbf{X})| \geq \frac{c_1}{N^{1/4}} \right) \leq \frac{c_2}{N^{1/4}}$$

holds, where  $\delta f(\mathbf{X})$  is the difference between the time average starting from  $\mathbf{X}$  and the ensemble average, and  $c_1, c_2$  are numerical constants. In the above equation the probability of the event  $|\delta f(\mathbf{X})| \geq c_1 N^{-1/4}$  is computed according to the microcanonical distribution.

The result, originally obtained by Khinchin for systems of non interacting particles, and then extended by Mazur and van der Linden [16], to systems of particles interacting through a short-range potential, can be summarized as follows: although the EH does not hold in a rigorous mathematical sense, it is still “physically” valid provided

- (a) we are interested only in suitable observables, with the shape (7);
- (b) we are a bit “tolerant” and accept that ergodicity holds  $O(N^{-1/4})$ , i.e., the difference between the time average and the ensemble average is  $O(N^{-1/4})$ ;
- (c) ergodicity can fail in regions whose probability is as small as  $O(N^{-1/4})$ , which vanishes for  $N \rightarrow \infty$ .

Khinchin’s result is a good example of what is usually called an *emergent property*: the dynamics has a marginal role, in the sense that in this context the crucial ingredient is the large number of particles  $N$ , being related to a technical aspect of probability theory, i.e., the *Law of large numbers* [17]. Of course the details of the dynamics can have also a rather important role, e.g., this is a quite clear in Ulam’s model discussed in Sect. 2.1. Basically in all the problems of non equilibrium statistical physics it is not possible to avoid a treatment in terms of the underlying dynamics, see, e.g., the many delicate topics in the celebrated Fermi-Pasta-Ulam problems [18].

### 2.4 On the Microcanonical Distribution

Let us open a short digression about the microcanonical distribution  $\rho_{micro}(\mathbf{X})$  which is constant if  $H(\mathbf{X}) \in [E, E + \Delta E]$  and zero otherwise.

Such an assumption sometimes is justified with the following argument: since we cannot be able to have a control of a complex system with a large number of components, we assume that probability distribution is uniform in the allowed region. The above reasoning is not convincing: consider the variable  $\mathbf{X}$  with a uniform probability density  $\rho_X(\mathbf{X})$  in a certain region, then the probability density  $\rho_Y(\mathbf{Y})$  of another variable  $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ , in general is not constant, therefore a uniform distribution, at variance with some folklore, has no special status among the possible distributions; in the following we will discuss again this point.

In our opinion the true physical good reason to adopt the microcanonical distribution is the following:  $\rho_{micro}(\mathbf{X})$  is a stationary solution of the Liouville’s equation. We can say that the dynamics, somehow, enters in the selection of the probability.

## 3 Critical Phenomena

The second case discussed by Wheeler can be understood in a similar way: in critical phenomena, the universal behavior near the critical point is explained by means of the Renormalization Group, which has a probabilistic interpretation in terms of generalized central limit theorems for non independent aleatory variables [19].

It is well known that, for the validity of the central limit theorem, two necessary conditions must be met: the random variables must have a finite variance and be uncorrelated (or weakly correlated). Assume that the random variables  $x_1, \dots, x_n$  are independent and identically distributed with  $p(x)$  having zero mean and  $\sigma < +\infty$ , the variable  $y_n = \frac{1}{\sigma\sqrt{n}}(x_1 + \dots + x_n)$  is distributed according to a normalized Gaussian distribution. In the case  $p(x) \sim |x|^{-\alpha}$ , with  $1 < \alpha < 3$ ,  $\sigma = +\infty$  and we have another limit theorem: the variable  $y_n = \frac{1}{n^\beta}(x_1 + \dots + x_n)$ , with  $\beta = 1/(\alpha - 1)$ , for  $n \gg 1$ , is distributed according to the so-called Lévy stable distribution  $\mathcal{L}_\alpha(x)$ ,

and at large  $|x|$  one has  $\mathcal{L}_\alpha(x) \sim |x|^{-\alpha}$ . As an explicit example, where it is possible to write  $\mathcal{L}_2(x)$ , we can cite the case of the Cauchy-Lorentz distribution  $P_\gamma(x) = \frac{\gamma}{\pi}(x^2 + \gamma^2)^{-1}$ , whose variance is infinite and in this case, for large  $n$ , the sample average  $\frac{1}{n}(x_1 + \dots + x_n)$  is not distributed according to a Gaussian, rather it follows the Cauchy-Lorentz distribution with exactly the same parameter  $\gamma$  as the individual variables. Thus, averaging Cauchy-Lorentz-distributed variables does not lead to a narrower distribution.

The ubiquitous occurrence of the Gaussian distribution in statistical physics is ultimately due to the fact that correlations are usually short-ranged. However, approaching the critical point that marks a second order phase transition, with varying a control parameter like the temperature, the microscopic degrees of freedom become correlated on larger and larger scales. As a result, very different physical systems manifest striking similarities in their nearly critical behavior. For instance, several physical quantities tend to diverge (or vanish) as a power law when approaching the phase transition point, and different systems are characterized by the very same set of critical exponents (the numbers that characterize the power-law divergence, or vanishing, of various physical quantities). All the systems sharing the same set of critical exponents are said to belong to the same *class of universality* [20]. In Wheeler’s perspective, the question might be cast in the following manner: How can very different physical systems “know” that they will share the same set of critical exponents? It is now clear that the universality of critical phenomena can be understood in terms of deviations from the central limit theorem, and a new class of limit theorems in probability theory is indeed required to properly describe the properties of a physical system near a second order phase transition.

The idea underlying critical phenomena is that the variables that provide a microscopic description of a given system become more and more correlated when approaching a phase transition, so that the conditions that guarantee that the macroscopic variables, obtained by averaging over the microscopic degrees of freedom, follow a Gaussian distribution (as a consequence of the central limit theorem) are violated. The renormalization group approach provides the technical tool to derive the asymptotic distributions. A prototypical model [21] that illustrates this increasingly strong correlation is described by a Hamiltonian defined on a set of  $2^N$  variables  $x_i$ , obeying the following hierarchical structure (see also Fig. 2):

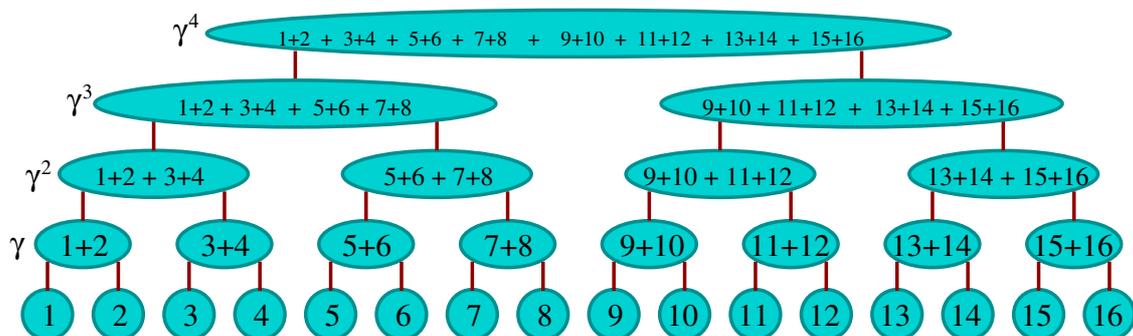


Fig. 2 Example of a hierarchical model with  $N = 4$

$$H_k(x_1, \dots, x_{2^k}) = H_{k-1}(x_1, \dots, x_{2^{k-1}}) + H_{k-1}(x_{2^{k-1}+1}, \dots, x_{2^k}) - \gamma^k \left( \sum_{i=1}^{2^k} \frac{x_i}{2} \right)^2, \tag{8}$$

with  $k = 1, \dots, N$ ,  $H_0 = 0$  and the coupling  $\gamma$  satisfying  $1 < \gamma < 2$  (for  $\gamma < 1$  the hierarchical interaction scales to zero and the system is trivial, for  $\gamma > 2$  the system is thermodynamically unstable). Let us assume for simplicity that the distribution of the individual  $x_i$  is Gaussian:

$$p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

with unit variance. In this case, starting from the Gibbs distribution of  $2^N$  variables,

$$e^{-\beta H_N(x_1, \dots, x_{2^N})} \prod_{i=1}^{2^N} p_0(x_i),$$

we can analytically proceed to regroup variables into blocks (lets call  $\mathcal{R}$  the action of this regrouping procedure on the distribution) and obtain a recursion relation for the distribution of the *block variables*

$$x^{(k)} = \sum_{i=1}^{2^k} \frac{x_i}{2^k}. \tag{9}$$

After  $k$  iterations one has

$$p_k(x) \equiv (\mathcal{R}^k p_0)(x) = \frac{1}{\sqrt{2\pi \sigma_k^2}} e^{-x^2/(2\sigma_k^2)},$$

with variance

$$\sigma_k^2 = \frac{1}{1 - 2\beta \sum_{\ell=1}^k \left(\frac{\gamma}{2}\right)^\ell}.$$

For  $k \rightarrow \infty$  the limiting distribution is Gaussian,

$$p_\infty(x) = \frac{1}{\sqrt{2\pi \sigma_\infty^2}} e^{-x^2/(2\sigma_\infty^2)},$$

with variance

$$\sigma_\infty^2 = \frac{1}{1 - \frac{2\beta\gamma}{2-\gamma}},$$

provided  $\beta < \beta_{crit} \equiv \frac{1}{\gamma} - \frac{1}{2}$ . Thus, in this case, the conditions required for validity of the central limit theorem hold at sufficiently high temperature: the variables are strongly correlated at short scale, but become weakly correlated at sufficiently large scales, ensuring that the asymptotic distribution is still Gaussian, with a suitable variance.

Quite interestingly, at  $\beta = \beta_{crit}$ , the variance diverges and the asymptotic distribution for the block variables (9) can no longer be Gaussian, signaling that the variables are correlated at all scales of the hierarchy. This is exactly what happens in a physical system close to a critical point. To obtain a well-defined limiting distribution, a different strategy to regroup the variables, and a different limit theorem, must be used. Let us introduce the new *block variables*

$$x^{(k)} = \sum_{i=1}^{2^k} \frac{x_i}{2^k \gamma^{-k/2}}, \quad (10)$$

different from Eq. (9). Again, the asymptotic distribution for these variables can be worked out analytically. The result is that for  $\sqrt{2} < \gamma < 2$  the distribution is Gaussian with variance

$$\sigma_\infty^2 = \frac{2 - \gamma}{2\gamma\beta_c}.$$

Even more interesting is the result for  $1 < \gamma < \sqrt{2}$ . Things become more intricate, and a simple solution can only be found for small  $\epsilon \equiv \sqrt{2} - \gamma > 0$ . In this case the asymptotic distribution of the block variables (10) is non Gaussian,

$$p_\infty(x) \approx A e^{-r^*(\epsilon)x^2/2 - u^*(\epsilon)x^4/4},$$

where  $A$  is a normalization constant and the parameters  $r^*(\epsilon)$  and  $u^*(\epsilon) > 0$  characterize the distribution and are called fixed-point couplings.

Thus, the hierarchical model (8) is rich enough to provide a realization of nontrivial limit theorems embodying the universal behavior of a system near a phase transition.

Let us conclude this section stressing the deep link between the renormalization group procedure and a generalization of the central limit theorem. There is an interesting class of stable distribution laws of random variables with the following property: let  $P_\lambda(x)$  be the probability of the variable  $x$  depending on the set of parameters  $\lambda$  (e.g., in a Gaussian distribution the  $\lambda$ 's amount to the mean value and the variance); denoting with  $*$  the convolution, we have that for the stable distribution law

$$(P_{\lambda_1} * P_{\lambda_2})(x) = P_\lambda(x),$$

where  $\lambda$  is an appropriate function of  $\lambda_1$  and  $\lambda_2$ . For instance, in the case of a Gaussian distribution  $\lambda = (m, \sigma^2)$  with  $m = m_1 + m_2$  and  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ . It is remarkable that the renormalization group can be seen as a convolution with the rule

$$P_{\lambda_n}(x) = P_{\lambda_{n-1}}(\alpha_{n-1}x) * P_{\lambda_{n-1}}(\alpha_{n-1}x),$$

and one has a fixed point as  $n \rightarrow \infty$  for a suitable choice of  $\lambda_n$  and  $\alpha_n$ .

The simplest Gaussian fixed point (for simplicity we take  $m = 0$ ) corresponds to  $\alpha_n = 1/\sqrt{2}$  and  $\sigma_n^2 = \sigma_{n-1}^2$ .

## 4 Discussion

The present Section is devoted to the discussion of a some technical, as well as conceptual, aspects of the probability in statistical physics which are often underestimated.

### 4.1 The Maximum Entropy Principle

We wish to add some remarks about information and the “it from bit” approach. The question here is whether statistical mechanics can be seen as a form of statistical inference. According to a radically anti-dynamical point of view, this is indeed the case: statistical mechanics is not a theory of objective physical reality, and the probabilities measure the “degree of truth” of a logical proposition. In this context, Jaynes [22] proposed the maximum entropy principle (MEP) as a general rule to infer the probability of a given event when only partial information is available. Let us briefly summarize the MEP approach: the mean values of  $M$  independent functions  $f_k(\mathbf{X})$  are known,

$$c_k = \langle f_k \rangle = \int f_k(\mathbf{X})\rho(\mathbf{X}) d\mathbf{X}, \quad \text{for } k = 1, \dots, M \tag{11}$$

then the MEP rule determines the probability density  $\rho(\mathbf{X})$  maximizing the entropy

$$\mathcal{H} = - \int \rho(\mathbf{X}) \log \rho(\mathbf{X}) d\mathbf{X},$$

with the constraints (11). Using the method of Lagrange multipliers, it is then easy to show that

$$\rho(\mathbf{X}) = \text{Const.} \exp \left[ \sum_{k=1}^M \lambda_k f_k(\mathbf{X}) \right],$$

where  $\lambda_1, \dots, \lambda_M$  depend on  $c_1, \dots, c_M$ . When applied to the statistical mechanics of systems with fixed number of particles and the only constraint of a given mean value  $E$  for the energy, the MEP recipe leads to the canonical distribution in a very simple way.

As far as we know this is the unique relevant success of MEP in physics. Up to now there is not any convincing evidence of the possibility to use the MEP to derive unknown results, e.g. in non equilibrium statistical mechanics [23]; in the following we will discuss the reason of this difficulty.

## 4.2 About Probability and Typicality

Let us briefly comment upon the fact that the approach in terms of ergodicity and that based on MEP reflect deep conceptual differences about how to consider probability. It is well known that there are several competing interpretations of the actual meaning of probability [24]. It is rather difficult to discuss in detail the different interpretations, here our interest is limited to the probability in statistical mechanics, in particular to the link between probability and real world. Our main aim is to investigate the following topic: what is the link between the probabilistic computations and the actual results obtained in laboratory experiments. Roughly speaking the different points of view about probability can be classified in two large classes: the subjective interpretation and the objective one.

According to the subjective interpretation probability is a degree of belief; one of the most influential supporter of such a point of view is Jaynes with his idea that the theoretical description of physical systems is governed by the degree of belief of the observer.

On the contrary, in the objective interpretation the probability of an event is determined by the physics of the systems and not by the lack of information of the observer. In particular ergodic theory, somehow, justifies a frequentist interpretation of probability, according to which it is possible to obtain an empirical notion of probability which is an objective property of the trajectory [5,7,25]. There is no universal agreement on this issue; for instance, Popper believed that probabilistic concepts are extraneous to a deterministic description of the world, while Einstein held the opposite point of view [26].

Let us try to give an answer to the following bold question: what is the link between the probabilistic computations (i.e., the averages over an ensemble) and the results obtained in laboratory experiments which, a fortiori, are conducted on a single realization (or sample) of the system under investigation? In a nutshell, following Boltzmann, we can invoke the notion of typicality, i.e., the fact that the outcome of an experiment on a macroscopic system takes a specific (typical) value overwhelmingly often [27–29]. In statistical mechanics typicality holds for  $N \gg 1$ . The concept of typicality is at the basis of the very possibility to have reproducibility of results in experiments (on macroscopic objects) or the possibility to have macroscopic laws.

From the limit theorems (as well as the typicality) one has that the true deep role of probability in statistical mechanics is the possibility to identify the mean value with the actual result for a unique (large) system [27–29].

The practical relevance of the typicality can be appreciated looking carefully at the most popular computational methods in statistical mechanics, i.e. molecular dynamics and Monte Carlo. The first technique is based on the assumption of the validity of the ergodicity. As already discussed, strictly speaking, such an assumption is surely wrong. In the Monte Carlo computations one uses suitable ergodic Markov chains, therefore the validity of the ergodicity is sure. On the other hand it is easy to realize that the success of the method cannot rely only on such a mathematical base. From Poincaré's theorem, and Kac's lemma on the mean recurrence time, one can understand [6] that for some observables in order to obtain the proper mean value from a time average, a

gigantic number of steps is necessary. Namely a time  $O(C^N)$  where  $C > 1$  and  $N$  is the number of degrees of freedom [30].

Therefore the true reason of the success cannot be the mathematical ergodicity, but the typicality and it is, somehow, related to the Khinchin’s results, i.e.

- the investigated system has a huge number of degrees of freedom:  $N \gg 1$ ;
- in a Monte Carlo computation, usually one computes only the average of a few observables of physical interest which involve many degrees of freedom and therefore it is not necessary to explore the whole phase space.

### 4.3 Is then the MEP a Cornucopia, or a Pandora’s Box?

We raise here two main objections. The first comes from the old good principle *Ex nihilo nihil* (Nothing from nothing).

In his nice book about statistical mechanics [31] Ma asks: “How many days in a year does it rain in Hsinchu (a city in northern Taiwan, commonly nicknamed *The Windy City*, because of its windy climate)?”. According to MEP the reply might be: “As there are two possibilities, to rain or not to rain, and we are completely ignorant about Hsinchu, it rains six months a year”. We share Ma’s opinion that the above answer is simply nonsense: it is not possible to infer something about a real phenomenon, only based on our ignorance.

But there is a more technical aspect that makes the MEP approach rather weak, namely the dependence of the results on the choice of the variables: the MEP gives different solutions if different variables are adopted to describe the same phenomenon. This is due to the fact that the “entropy”

$$\mathcal{H}_X = - \int \rho_X(\mathbf{X}) \log \rho_X(\mathbf{X}) \, d\mathbf{X}$$

is not an intrinsic quantity but it depends on the variable  $X$  used for the state of the system. Using a different parametrization, i.e., the coordinates given by an invertible map  $\mathbf{Y} = \mathbf{f}(\mathbf{X})$ , the entropy of the same phenomenon would be

$$\mathcal{H}_Y = - \int \rho_Y(\mathbf{Y}) \log \rho_Y(\mathbf{Y}) \, d\mathbf{Y} = \mathcal{H}_X + \int \rho_X(\mathbf{X}) \log |D\mathbf{f}(\mathbf{X})| \, d\mathbf{X},$$

where  $|D\mathbf{f}(\mathbf{X})|$  is the determinant of the Jacobian matrix, measuring how the elementary volume changes with the change of variables.

Let us now reconsider the fact that using MEP is easy to obtain the canonical distribution. The supporters of MEP consider this result an important success of the method. On the other hand there is clearly a negative aspect: the correct result is obtained only provided we make use of the canonical variables, i.e., positions and momenta of the particles. The choice of the specific canonical variables is not unique, but different choices are related to one another by canonical transformations that preserve the Hamiltonian structure of the equations of motion and enjoy the property that the corresponding determinant of the Jacobian matrix equals unity. On the con-

trary using different (non canonical) variables, instead of the Const.  $e^{-\beta H}$  we have  $g(\mathbf{Y})e^{-\beta H(\mathbf{Y})}$  where the shape of the function  $g(\ )$  depends on the chosen variables.

## 5 Final Remarks

Let us conclude the paper with some general comments and remarks.

Wheeler’s slogan “There is no law except the law that there is no law” sometimes can be interpreted in a precise way: In probability theory there exist limit theorems (law of large numbers, central limit theorems) which allow for the understanding of the behavior of physical systems with many degrees of freedom. Some words of caution are needed here. Sometime it is not possible to avoid some details from fundamental theories, for instance, in the case of the Boltzmann–Gibbs canonical distribution. In such a case the basic assumption is the validity of the microcanonical distribution which can be justified by a dynamical argument, namely the Liouville theorem and the assumption of (a weak form of) ergodicity.

It is dangerous to believe too much in *general principles* (like MEP), whose results, sometimes, can be used to simplify a description *a posteriori*, i.e., only after a deeper understanding of a given piece of physical reality by means of a more robust mathematical approach. An example of the above warning is provided by Ulam’s conjecture on Boltzmann’s law, which can be obtained with a stochastic rule in the binary collision process. A careful analysis shows how the correct result is obtained only with a precise collision rule, and therefore it cannot be considered just the gift of a general regulatory principle.

To conclude, we stress again the basic role of the limit theorems in statistical mechanics. The law of large numbers and the typicality are the two relevant ingredients which allow for a conceptual (and technical) link between the use of probability and the observations in a unique (large) system.

**Acknowledgements** We wish to thank M. Falcioni and A. Puglisi for many useful comments.

## References

1. Wheeler, J.A.: On recognizing “law without law”, oersted medal response at the Joint APS-AAPT meeting, New York, 25 January 1983. *Am. J. Phys.* **51**, 398 (1983)
2. Wheeler, J.A.: Recent thinking about the nature of the physical world—it from bit. *Ann. N. Y. Acad. Sci.* **655**, 349 (1992)
3. Deutsch, D.: On Wheeler notion of law without law in physics. *Found. Phys.* **16**, 565 (1986)
4. Emch, G., Liu, C.: *The Logic of Thermostatistical Physics*. Springer, New York (2002)
5. Chibbaro, S., Rondoni, L., Vulpiani, A.: *Reductionism, Emergence and Levels of Reality*. Springer, New York (2014)
6. Castiglione, P., Falcioni, M., Lesne, A., Vulpiani, A.: *Chaos and Coarse Graining in Statistical Mechanics*. Cambridge University Press, Cambridge (2008)
7. Cercignani, C.: *Ludwig Boltzmann: The Man Who Trusted Atoms*. Oxford Univ Press, Oxford (1998)
8. Blackwell, D., Mauldin, R.D.: Ulam’s redistribution of energy problem. *Lett. Math. Phys.* **60**, 149 (1985)
9. Bertrand, J.: *Calcul des probabilités*. Gauthier-Villars, Paris (1889)

10. Puglisi, A., Loreto, V., Marconi, U.M.B., Petri, A., Vulpiani, A.: Clustering and non-gaussian behavior in granular matter. *Phys. Rev. Lett.* **81**, 3848 (1998)
11. Nagel, E.: *The Structure of Science*. Hackett Publ. Comp, Indianapolis (1979)
12. Cerino, L., Puglisi, A., Vulpiani, A.: A consistent description of fluctuations requires negative temperatures. *J. Stat. Mech.* **2015**, P12002 (2015)
13. Dumas, H.S.: *The KAM Story*. World Scientific, Hackensack (2014)
14. Ciccotti, G., Hoover, W.G. (eds.): *Molecular Dynamics Simulations of Statistical Mechanical Systems*. North-Holland, Amsterdam (1986)
15. Khinchin, A.I.: *Mathematical Foundations of Statistical Mechanics*. Dover Publ, New York (1949)
16. Mazur, P., van der Linden, J.: Asymptotic form of the structure function for real systems. *J. Math. Phys.* **4**, 271 (1963)
17. Koralov, L.B., Sinai, Y.G.: *Theory of Probability and Random Processes*. Springer, New York (2007)
18. Gallavotti, G. (ed.): *The Fermi-Pasta-Ulam Problem: A Status Report*. Springer, New York (2007)
19. Jona-Lasinio, G.: Renormalization group and probability theory. *Phys. Rep.* **352**, 439 (2001)
20. Di Castro, C., Raimondi, R.: *Statistical Mechanics and Applications in Condensed Matter*. Cambridge University Press, Cambridge (2015)
21. Bleher, P.M., Sinai, J.G.: Investigation of the critical point in models of the type of Dyson's hierarchical models. *Commun. Math. Phys.* **33**, 23 (1973)
22. Jaynes, E.T.: Foundations of probability theory and statistical mechanics. In: Bunge, M. (ed.) *Delaware Seminar in the Foundations of Physics*, p. 77. Springer, Berlin (1967)
23. Auletta, G., Rondoni, L., Vulpiani, A.: On the relevance of the maximum entropy principle in non-equilibrium statistical mechanics. *Eur. Phys. J. Spec. Top.* **226**, 2327 (2017)
24. Kuzemskij, A.L.: Probability, information and statistical mechanics. *Int. J. Mod. Phys.* **55**, 1378 (2015)
25. von Plato, J.: *Creating Modern Probability*. Cambridge University Press, Cambridge (1994)
26. Popper, K.R.: *The Logic of Scientific Discovery*. Routledge, London (2002)
27. Goldstein, S.: Typicality and notions of probability in physics. In: Ben-Menahem, Y., Hemmo, M. (eds.) *Probability in Physics*, pp. 59–71. Springer, New York (2012)
28. Cerino, L., Cecconi, F., Cencini, M., Vulpiani, A.: The role of the number of degrees of freedom and chaos in macroscopic irreversibility. *Physica A* **442**, 486–497 (2016)
29. Zanghì, N.: I fondamenti concettuali dell'approccio statistico in fisica. In: Allori, V., Dorato, M., Laudisa, F., Zanghì, N. (eds.) *La natura delle cose: Introduzione ai fondamenti e alla filosofia della fisica*, Carocci, pp. 202–247 (2005)
30. Kac, M.: On the notion of recurrence in discrete stochastic processes. *Bull. Am. Math. Soc.* **53**, 1002 (1947)
31. Ma, S.K.: *Statistical Mechanics*. World Scientific, Philadelphia (1985)