

Entropy of Continuous Mixtures and the Measure Problem

Pablo Maynar

Física Teórica, Universidad de Sevilla, Apartado de Correos 1065, E-41080, Sevilla, Spain

Emmanuel Trizac

LPTMS (CNRS UMR 8626), Université Paris-Sud, Orsay Cedex, F-91405, France

(Received 12 October 2010; revised manuscript received 11 March 2011; published 21 April 2011)

In its continuous version, the entropy functional measuring the information content of a given probability density may be plagued by a “measure” problem that results from improper weighting of phase space. This issue is addressed considering a generic collision process whereby a large number of particles or agents randomly and repeatedly interact in pairs, with prescribed conservation law(s). We find a sufficient condition under which the stationary single-particle distribution function maximizes an entropylike functional, that is free of the measure problem. This condition amounts to a factorization property of the Jacobian associated with the binary collision law, from which the proper weighting of phase space directly follows.

DOI: 10.1103/PhysRevLett.106.160603

PACS numbers: 05.20.Dd, 02.50.Ng

In information theory, the definition of the entropy of a continuous probability distribution depends on the identification of a relevant prior, or weighting function [1,2], that can prove elusive. To illustrate this point and motivate our approach, let us consider an ensemble of particles (each indexed by integer i) that can exchange some positive quantity x so that $\sum_i x_i$ is fixed: two particles i and j chosen at random interact so that $x_i \rightarrow x_i + \eta$ and $x_j \rightarrow x_j - \eta$, provided both quantities remain positive. Here, η is a fixed small increment, or can be drawn from a prescribed distribution. Such a model has appeared in different settings: in the context of mass transport models, x stands for the mass of the particle [3]; it can also be the position of a composite object in exclusion processes [4], the volume of some colloidal aggregate [5], the size of a self-assembled polymer [6], the wealth of an agent in a simplistic econophysics framework [7], or an auxiliary quantity used for algorithmic purposes, in particular, the generation of pseudorandom numbers [8]. Upon iterating the previous “collision” rule, it can be shown that the x distribution reaches the simple stationary probability density function $p_X(x) = \exp(-x)$, fixing for convenience the mean x to unity [3]. Following the early work of Shannon [9], this result *seems* to be readily recovered by maximizing the information measure—or differential entropy—of the distribution

$$S_{\text{Shannon}} = - \int p_X(x) \log[p_X(x)] dx \quad (1)$$

under the constraint that $\int p_X dx = \int x p_X dx = 1$ [10]. On the other hand, it is clear that the process could be equally well described by another quantity y (say, the radius of a colloid instead of its volume), with a corresponding probability density p_Y such that $p_X(x) dx = p_Y(y) dy$. However, the formulation (1) is not invariant under change of variable, so that a different and inconsistent distribution would

be found by maximizing $-\int p_Y \log p_Y$, even after taking into account for constraints appropriately. We will refer to this latent deficiency, already noted in [9], as the “measure problem.” In addition, although the pathological nature of Eq. (1) is made evident by a change of variable, it can also be inferred from its dimensional inconsistency. We conclude that recovering the correct (in our example exponential) distribution from maximizing (1) is coincidental, and that Eq. (1) does not provide an admissible information measure. We emphasize that Shannon faced the measure problem [11], and concluded that the entropy of a continuous distribution is not an absolute measure, but is relative to the coordinate system. Such a point of view is not acceptable: the entropy should not have an absolute status for discrete probabilities, and a relative one for continuous cases.

The mechanism for recovering an absolute information measure that is unaffected by a parameter change, is clear when the continuous limit is carefully taken from the situation described by a discrete probability set $\{p_\alpha\}$, where the entropy reads $-\sum_\alpha p_\alpha \log p_\alpha$ [2,9,12]. In doing so, it is necessary to introduce the density of points $m_X(x)$ and one obtains [1]

$$S = - \int p_X(x) \log[\Lambda_X(x) p_X(x)] dx. \quad (2)$$

In the above expression, that seems to have been first derived and commented by Jaynes [1,13], the quantity Λ_X can be viewed as a weighting function, and it is related to the density by $\Lambda_X(x) = 1/m_X(x)$. This x -dependent function indicates how the space of dynamical variables is resolved [2]: the larger the density m_X , the better the resolution, which corresponds to a smaller Λ_X . Since the densities m transform under change of variable as the probability densities p do, the coordinate dependence

of Λ cures the measure problem. It is therefore essential to understand what this dependence is, a problem that is quite often overlooked in the literature [5,6,14] and that Jaynes—somewhat ironically—attributes to the fact that “one could not think of anything else to do” [15]. Consequently, if the density m_X can be extracted from our knowledge of x -space sampling, the measure problem is solved. This case is that of a “quenched” x distribution. There are, nevertheless, situations where this knowledge is not *a priori* available, but is encoded in the dynamics of the system (“annealed” x distribution), so that m_X is selected by the underlying dynamical rules. Our goal here is to understand that connection in an annealed context, in order to set up a clear prescription for writing the relevant entropy.

We are now in a position to state the problem more precisely. We are interested in a population of a large number N of particles where a given property x_i (mass, velocity, length, color, income, etc.) is attached to each particle i . These particles undergo repeatedly binary “collisions” where pairs selected at random interact such that $(x_i, x_j) \rightarrow (x'_i, x'_j)$. An important point is that we assume the existence of a conservation law

$$\mathcal{C}(x_i) + \mathcal{C}(x_j) = \mathcal{C}(x'_i) + \mathcal{C}(x'_j), \quad (3)$$

where \mathcal{C} is a given function. We shall leave ergodicity issues aside, and consider that the functions $x'_i(x_i, x_j)$, $x'_j(x_i, x_j)$, that are not specified, are sufficiently mixing to ensure that all accessible phase space is sampled (in general, nonuniformly). The objective is to answer the following question. \mathcal{Q} : Can we maximize a functional of the form (2), under the appropriate constraints that $\int p_X(x)dx$ and $\int \mathcal{C}(x)p_X(x)dx$ are fixed, to obtain the steady-state probability distribution $p_X^{st}(x)$, if it exists? If so, simple calculus shows that the latter distribution is of the form

$$p_X^{st}(x) = \alpha \Lambda_X^{-1}(x) \exp[-\beta \mathcal{C}(x)] \quad (4)$$

where α and β are irrelevant Lagrange multipliers. The ensuing problem is then to understand what specifies the weighting function Λ_X . Indeed, knowing that \mathcal{Q} can be answered affirmatively is of little interest if one does not know the corresponding weighting function $\Lambda_X(x)$.

The collision law considered may violate detailed balance, and it may involve an additional stochastic parameter η , as for instance in the simple example introduced in [8], that we mention as a warm-up exercise:

$$\begin{cases} x'_1 \\ x'_2 \end{cases} = \begin{cases} \eta(x_1 + x_2)/\sqrt{2} \\ \eta(x_1 - x_2)/\sqrt{2} \end{cases}, \quad (5)$$

where η equiprobably takes values ± 1 . Likewise, randomness is necessarily introduced for colliding hard bodies, as a remnant of the impact parameter in a description that only considers the velocity degrees of freedom, as routinely done in some Monte Carlo simulation techniques [16].

It should be clear from the outset that the conserved quantity is in general not exponentially distributed, as our simplistic introductory example might lead to believe. Indeed, considering Eq. (5) that conserves “energy” $e \equiv x^2$, i.e., $\mathcal{C}(x) = x^2$, it appears that $p_E^{st}(e) \propto \exp(-\beta e)/\sqrt{e}$ in the steady state [17], where β is some inverse temperature. A naive application of Eq. (1), on the other hand, leads to the incorrect result $p_E^{st}(e) \propto \exp(-\beta e)$. This means that here $\Lambda_E(e) \propto \sqrt{e}$, and we learn on this simple example that the conservation law is not sufficient, in general, to obtain the relevant Λ . This key quantity is encoded in the transformation law $(x_i, x_j) \rightarrow (x'_i, x'_j)$, in a way that we now bring to the fore.

To get an idea of the connection ($\Lambda \leftrightarrow$ dynamics), we first restrict to the subclass of processes that fulfill detailed balance. The corresponding single-particle distribution obeys then

$$p_X^{st}(x_1)p_X^{st}(x_2)dx_1dx_2 = p_X^{st}(x'_1)p_X^{st}(x'_2)dx'_1dx'_2, \quad (6)$$

where, due to the mean-field-like sampling procedure with randomly chosen pairs, the two-particle probability distribution factorizes for large N into a product of single-particle distributions (a more technical proof will be outlined below). On the other hand, assuming that \mathcal{Q} can be answered positively, the stationary single-particle distribution p_X^{st} is constrained to be of the form (4). Then, from Eqs. (4) and (6) and the conservation law, Eq. (3), we find that the Jacobian \mathcal{J} of the transformation $(x_1, x_2) \rightarrow (x'_1, x'_2)$, admits a factorized form

$$\mathcal{J}(x_1, x_2) \equiv \left| \det \frac{\partial(x'_1, x'_2)}{\partial(x_1, x_2)} \right| = \frac{\Lambda_X(x'_1)\Lambda_X(x'_2)}{\Lambda_X(x_1)\Lambda_X(x_2)}. \quad (7)$$

We emphasize here that the Jacobian is defined for a given value of the stochasticity parameter η : x'_1 and x'_2 are functions of x_1, x_2 , and η .

We now arrive at our main part and we will show below that if the factorization property (7) of \mathcal{J} holds (without any other restriction as, for example, detailed balance) then the stationary distribution function p_X^{st} is of the form (4), and hence we are able to answer affirmatively to question \mathcal{Q} . In addition, the relevant weighting function Λ_X can then be directly read from (7). This is interesting from an operational point of view, since the Jacobian directly follows from the knowledge of the collision law, which is an input of the model. As an illustration, we return to the toy model of Eq. (5), recast in the conserved variable $e \equiv x^2$. We have $\mathcal{J}(e_1, e_2) = [e'_1 e'_2 / (e_1 e_2)]^{1/2}$, which is of the form (7), with $\Lambda_E(e) \propto \sqrt{e}$. This immediately leads to the correct distribution $p_E^{st}(e) \propto \exp(-\beta e)/\sqrt{e}$.

We now proceed with our general proof, that starts with assuming property (7) for \mathcal{J} , and that involves the following three steps. (a) We introduce a new set of variables, under the mild assumption that Λ_X in (7) is nonvanishing. Indeed, with the function of x $z(x) = \int^x dx'/\Lambda_X(x')$, the Jacobian of the collision law becomes unity ($dz'_1 dz'_2 = dz_1 dz_2$),

which simplifies the kinetic theory description. (b) Although our aim is to derive the stationary single-particle distribution function ρ_Z , working at N -body level with the phase space density $\rho_N(\Gamma, t)$, where $\Gamma \equiv (z_1, \dots, z_N)$, turns out to be a convenient detour. This distribution obeys the following evolution equation [18,19]:

$$\partial_t \rho_N(\Gamma, t) = \sum_{i < j}^N \int d\eta w(\eta) [b_{ij}^{(\eta)} - 1] \rho_N(\Gamma, t), \quad (8)$$

where the random variable η with distribution w enters the collision law (see above), that can be described by the inverse collision operator $b_{ij}^{(\eta)}$. This operator acts on the distribution on its right by replacing the arguments z_i and z_j by their precollisional values z_i^* and z_j^* :

$$b_{12}^{(\eta)} \rho_N(\Gamma, t) = \rho_N(z_1^*, z_2^*, z_3, \dots, z_N, t) \quad (9)$$

with $(z_{i,j}^*)' = z_{i,j}$. The present description in terms of z quantities is also endowed with a conservation law, that we write here—modulo a slight abuse of notation—with the same function \mathcal{C} as in Eq. (3): $\sum_i \mathcal{C}(z_i) = C$. It is then straightforward to see that the distribution $\rho_N^{st} \propto \delta(C - \sum_i \mathcal{C}(z_i))$ (with proper normalization) provides a stationary solution to Eq. (8). The corresponding single-particle distribution function follows from computing the first marginal $\rho_X^{st}(z_1) \propto \int \rho_N^{st} dz_2 \dots dz_N$. The argument is akin to that put forward to construct the canonical ensemble from the microcanonical distribution [20], and leads to $p_Z^{st}(z) \propto \exp[-\beta \mathcal{C}(z)]$. (c) The last important step in the argument is to show that the N -body measure ρ_N^{st} is attractive, at long times, for arbitrary initial conditions sharing the same value of C . For this purpose, we borrow a technique introduced in [21] and consider an arbitrary strictly convex positive function $h(x)$ from which we construct

$$H(t) = \int d\Gamma \rho_N^{st}(\Gamma) h[\rho_N(\Gamma, t)]. \quad (10)$$

The evolution equation (8) implies

$$\begin{aligned} \frac{dH}{dt} &= \frac{N(N-1)}{2} \int d\Gamma d\eta \rho_N^{st}(\Gamma) w(\eta) \{h'[\rho_N(\Gamma, t)] \\ &\quad \times [\rho_N(b_{12}^{(\eta)} \Gamma, t) - \rho_N(\Gamma, t)] + h[\rho_N(\Gamma, t)] \\ &\quad - h[\rho_N(b_{12}^{(\eta)} \Gamma, t)]\}, \end{aligned} \quad (11)$$

where we have used the invariance under permutation of particle indices and that

$$\int d\Gamma d\eta \rho_N^{st}(\Gamma) w(\eta) \{h[\rho_N(\Gamma, t)] - h[\rho_N(b_{12}^{(\eta)} \Gamma, t)]\} = 0. \quad (12)$$

From the convexity of h , we have that H is a nonincreasing function of time. Since H is in addition bounded from below, we have established that it converges at long times to a constant [22]. Moreover, as the curly bracket of Eq. (11) only vanishes when $\rho_N(\Gamma, t) = \rho_N(b_{12}^{(\eta)} \Gamma, t)$, we conclude from our ergodicity assumption that all initial

phase space densities for which the conserved quantity is strictly equal to C evolve towards ρ_N^{st} . This is a flat and finite measure on the ensemble defined by $\sum_i \mathcal{C}(z_i) = C$, and can be seen as a generalized microcanonical density. A similar property also applies to the first marginal $\rho_Z(z, t)$ that is then attracted to $p_Z^{st} \propto \exp[-\beta \mathcal{C}(z)]$. Returning to the original variable x in which the problem was formulated, and bearing in mind that $dz/dx = \Lambda_X^{-1}(x)$, this yields the desired result that the stationary distribution $p_X^{st}(x)$ is of the form (4). Incidentally, we also obtain here that the 2-body distribution $p_{2,X}^{st}(x_1, x_2)$ factorizes in the steady state, in the product $p_X^{st}(x_1) p_X^{st}(x_2)$, as mentioned above.

So far, we have shown that under the assumption (7), the steady-state distribution p_X^{st} can be found by minimizing a functional of the form (2), with a known weighting function Λ_X , directly read from (7). This was illustrated by the toy dynamics (5), but our introductory example also may be understood in that framework: with the dynamics defined by $(x_i, x_j) \rightarrow (x_i + \eta, x_j - \eta)$, for which $\mathcal{C}(x) = x$, we simply have $\mathcal{J}(x_1, x_2) = 1$, hence $\Lambda_X = 1$ and $p_X^{st}(x) \propto \exp(-\alpha x)$.

To complete the analysis, three remarks are in order. First, while we have restricted to the scalar case for the sake of simplicity, x can equally be a vectorial quantity. Second, Eq. (7), when it applies, does not define a unique function Λ_X . Indeed, consider two candidates obeying

$$\frac{\Lambda_X(x'_1) \Lambda_X(x'_2)}{\Lambda_X(x_1) \Lambda_X(x_2)} = \frac{\tilde{\Lambda}_X(x'_1) \tilde{\Lambda}_X(x'_2)}{\tilde{\Lambda}_X(x_1) \tilde{\Lambda}_X(x_2)} \quad (13)$$

for all x_1, x_2 , and η . Then, $\log(\Lambda_X/\tilde{\Lambda}_X)$ is a collisional invariant. Assuming that there is no “hidden” conservation law, we have $\log(\Lambda_X(x)/\tilde{\Lambda}_X(x)) = a + b\mathcal{C}(x)$, where a and b are arbitrary constants. So, a candidate weighting function defined through (7), is prescribed up to a function $\exp(a + b\mathcal{C}(x))$. Such a freedom in the choice of Λ only shifts the functional (2) by the constant $a + b\langle \mathcal{C} \rangle$. The final result for p_X^{st} is hence not affected by the choice made for Λ_X .

Third, it seems worthwhile to provide a more intuitive understanding of the fact that if the Jacobian $(x_1, x_2) \rightarrow (x'_1, x'_2)$ fulfills Eq. (7), then the corresponding weighting function is proportional to the Λ_X appearing in (7). For any pair (x_1, x_2) , (7) implies that the respective uncertainties δx_1 and δx_2 , will be affected by the collision such that $\Lambda_X(x_1)^{-1} \delta x_1 \Lambda_X(x_2)^{-1} \delta x_2 = \Lambda_X(x'_1)^{-1} \delta x'_1 \Lambda_X(x'_2)^{-1} \delta x'_2$. At long times, the system reaches a state where the solution to the above constraint is simply $\delta x/\Lambda_X(x) = \text{cst}$, so that at one body level, the space of dynamical variables is resolved with an x -dependent precision $\delta x \propto \Lambda_X(x)$. This allows us to view Λ_X as an x -dependent volume in the space of dynamical variable, that quantifies the “grain-ing” with which the space is resolved. Alternatively, this argument shows that the density of points generated in x space verifies $m_X(x) \propto 1/\delta x \propto 1/\Lambda_X(x)$, as already mentioned.

To summarize, we have studied a class of problems encountered in different contexts, such as soft matter where a mixture of polydisperse hard spheres [5,24,25], hard rods [26], or ring polymers [6] have been shown to exhibit a condensation in real space, stochastic mass transport models [3], or in mathematical literature where the Kac walk [21] is an important kinetic theory toy model for studying the propagation of chaos and rate of equilibration [27]. Specifically, our goal here was to analyze under which conditions the steady-state distribution p_X^{st} obtained by iterating a generic collision process with conservation law [Eq. (3)] could equivalently be obtained from a maximum entropy argument by extremalizing a given functional of the type (2). We have found that this is the case if the Jacobian of the collision law $(x_1, x_2) \rightarrow (x'_1, x'_2)$, can be written as in (7), from which the relevant weighting function $\Lambda_X(x)$ can be extracted, which provides simply p_X^{st} . This is, for example, the case of Refs. [5–8,24–26]. The connection thereby established is free of the so-called measure problem, that plagues a naive writing of the entropy functional as in (1), an expression first proposed by Shannon, and that propagated in a significant fraction of the literature. Our analysis, in other words, provides the correct prior Λ_X^{-1} that should be considered, see Eq. (2). A key point is that the configurations allowed by the conservation law(s) are in general sampled nonuniformly. This nonuniformity, encoded in the x dependence of Λ_X , that gives different weights to different points in x space, is the feature ensuring that the information measure considered is absolute, and does not depend on the parametrization chosen. We finally note that our approach—which includes multiple conservation laws—can be generalized to more complex collisional processes, involving more than two bodies, or in which the collision frequency $\omega(x_1, x_2)$, chosen constant here for the sake of simplicity, actually depends on the pair considered, as long as $\omega(x_1, x_2) = \omega(x'_1, x'_2)$.

[1] E. T. Jaynes, in *Information Theory and Statistical Mechanics*, Brandeis Lectures in Theoretical Physics (Brandeis Summer Institute in Theoretical Physics, Waltham, MA, 1962), Vol. 3, p. 181.
 [2] R. Balian, *From Microphysics to Macrophysics* (Springer-Verlag, Berlin, 2006).
 [3] S. N. Majumdar, in *Exact Methods in Low-dimensional Statistical Physics and Quantum Computing*, Proceedings of the Les Houches Summer School (Oxford University Press, New York, 2010); arXiv:0904.4097.
 [4] T. M. Liggett, *Interacting Particle Systems* (Springer-Verlag, New York, 1991).
 [5] J. Zhang, R. Blaak, E. Trizac, J. A. Cuesta, and D. Frenkel, *J. Chem. Phys.* **110**, 5318 (1999).
 [6] J. A. Cuesta and R. P. Sear, *Europhys. Lett.* **55**, 451 (2001); *Phys. Rev. E* **65**, 0314061 (2002).

[7] A. Dragulescu and V. Yakovenko, *Physica (Amsterdam)* **299A**, 213 (2001); *Eur. Phys. J. B* **17**, 723 (2000).
 [8] J. F. Fernández and C. Criado, *Phys. Rev. E* **60**, 3361 (1999).
 [9] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948); **27**, 623 (1948).
 [10] When quoting Shannon's seminal work [9], we do not refer to his well-known derivation of the entropy or information content of a discrete probability set, but to his generalization to continuous distributions (Pt. III, Sec. 20 of Ref [9]), that suffers from the measure problem discussed here. In addition, the example discussed here with $p_X = e^{-x}$ can be found in Sec. 20.7 of that paper.
 [11] See Pt. III.20.8 of Ref. [9].
 [12] T. M. Cover and J. M. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 2006), 2nd ed.
 [13] We note that Eq. (2) is reminiscent of a Kullback-Leibler distance (also called relative entropy) [12], with the difference that $m_X \propto 1/\Lambda_X$ is not a probability density.
 [14] See, e.g., J. J. Salacuse and G. Stell, *J. Chem. Phys.* **77**, 3714 (1982); M. R. Stapleton *et al.*, *J. Phys. Chem.* **92**, 4788 (1988); R. M. L. Evans, D. J. Fairhurst, and W. C. K. Poon, *Phys. Rev. Lett.* **81**, 1326 (1998).
 [15] See Ref. [1], p. 203. There is consequently a rather large body of literature that erroneously considers the measure problem—absent in the discrete case—as intrinsic to the continuous formulation, leaving aside the textbooks that do not mention the issue.
 [16] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon, Oxford, 1994).
 [17] The process defined here leads to a Gaussian distribution $p_X(x)$ [8]. The authors therefore proposed the method as a Gaussian random number generator. M. I. J. Probert, *Phys. Rev. E* **63**, 058701 (2001) showed the practical limitations of the initial algorithm (undesired correlations), which led the authors to put forward an improved version [J. F. Fernández and C. Criado, *Phys. Rev. E* **63**, 058702 (2001)].
 [18] Here, time t counts the number of collisions that occurred. The problem can be recast in a continuous t form.
 [19] P. Résibois and M. de Leener, *Classical Kinetic Theory of Fluids* (John Wiley, New York, 1977).
 [20] S. K. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985).
 [21] M. Kac, *Probability and Related Topics in Physical Sciences*, Lectures in Applied Mathematics Vol. 1A (American Mathematical Society, Providence, 1957).
 [22] Note that the fact that H is nonincreasing does not provide an H theorem in the usual sense [20,23], since our function depends on the final distribution. Additionally, because of the lack of detailed balance (which translates into $z_i^* \neq z_i'$), it is not possible to obtain directly an H functional at the single-particle level.
 [23] L. E. Reichl, *A Modern Course in Statistical Physics* (Wiley VCH, Weinheim, 2004), 2nd ed.
 [24] R. Blaak, *J. Chem. Phys.* **112**, 9041 (2000).
 [25] R. Blaak and J. A. Cuesta, *J. Chem. Phys.* **115**, 963 (2001).
 [26] M. R. Evans, S. N. Majumdar, I. Pagonabarraga, and E. Trizac, *J. Chem. Phys.* **132**, 014102 (2010).
 [27] E. Carlen, M. C. Carvalho, M. Loss, J. Le Roux, and C. Villani, *Kinet. Relat. Models* **3**, 85 (2010).