entropy

MDPI

Article

Thermodynamics Beyond Molecules: Statistical Thermodynamics of Probability Distributions

Themis Matsoukas

Department of Chemical Engineering, Pennsylvania State University, University Park, PA 16802, USA; txm11@psu.edu; Tel.: +1-814-863-2002

Received: 26 July 2019; Accepted: 11 September 2019; Published: 13 September 2019

Abstract: Statistical thermodynamics has a universal appeal that extends beyond molecular systems, and yet, as its tools are being transplanted to fields outside physics, the fundamental question, what is thermodynamics, has remained unanswered. We answer this question here. Generalized statistical thermodynamics is a variational calculus of probability distributions. It is independent of physical hypotheses but provides the means to incorporate our knowledge, assumptions and physical models about a stochastic processes that gives rise to the probability in question. We derive the familiar calculus of thermodynamics via a probabilistic argument that makes no reference to physics. At the heart of the theory is a space of distributions and a special functional that assigns probabilities to this space. The maximization of this functional generates the mathematical network of thermodynamic relationship. We obtain statistical mechanics as a special case and make contact with Information Theory and Bayesian inference.

Keywords: statistical thermodynamics; statistical mechanics; biased sampling; most probable distribution

1. Introduction

What is thermodynamics? The question, so central to physics, has been asked numerous times and has been given nearly as many different answers. To quote just a few: thermodynamics is "the branch of science concerned with the relations between heat and other forms of energy involved in physical and chemical processes" [1]; "the study of the restrictions on the possible properties of matter that follow from the symmetry properties of the fundamental laws of physics" [2]; "concerned with the relationships between certain macroscopic properties of a system in equilibrium" [3]; "a phenomenological theory of matter" [4]. Such statements, while strictly true, focus on aspects that are far too narrow to converge to a definition of sufficient generality as to *what* to call thermodynamics or *how* to carry it outside physics. And yet, since Gibbs [5], Shannon [6] and Jaynes [7] drew quantitative connections between entropy and probability distributions, thermodynamics has been spreading to new fields. The tools of statistical thermodynamics are now used in network theory [8], ecology [9], epidemics [10], neuroscience [11], financial markets [12], and in the study of complexity in general. What motivates the impulse to apply thermodynamics to such vastly diverse problems? Is thermodynamics even applicable outside classical or quantum mechanical systems? And if so, what is the scope of its applicability?

Here we answer these fundamental questions: Statistical thermodynamics is variational calculus applied to probability distributions and by extension to stochastic processes in general; it is independent of physical hypotheses but provides the means to incorporate our knowledge and model assumptions about the particular problem. The fundamental ensemble is a space of probability distributions sampled via a bias functional. The maximization of this functional expresses a distribution—any

distribution—via a set of parameters (microcanonical partition function, canonical partition function and generalized temperature) that are connected through a set of mathematical relationships that we recognize as the familiar equations of thermodynamic. Entropy and the second law have simple interpretations in this theory. We obtain statistical mechanics as a special case and make contact with Information Theory and Bayesian inference.

2. The Calculus of Statistical Thermodynamics

Before we derive a theory of generalized thermodynamics we review the key elements of the standard thermodynamic calculus. The central quantity of interest in statistical thermodynamics is the probability of microstate. For a system of N particles in volume V and temperature T this probability is given by the exponential (canonical) distribution,

$$\text{Prob(microstate } i) = \frac{e^{-\beta E_i}}{Q}, \tag{1}$$

where Q is the canonical partition function, E_i is the energy of microstate, $\beta = 1/k_B T$ and k_B is Boltzmann's constant. The corresponding probability to find the system in a microstate with energy E is obtained by summing all microstates with fixed energy E and is given by

$$\text{Prob}(E) = \Omega \frac{e^{-\beta E}}{Q}, \tag{2}$$

where Ω is the microcanonical partition function, also equal to the number of microstates with energy E, volume V and number of particles N. The mean energy $\bar{E}$ and the parameters Ω, Q and β that appear in Equation (2) are interrelated:

$$\log \Omega = \beta \bar{E} + \log Q, \tag{3}$$

$$\beta = \frac{\partial \log \Omega}{\partial \bar{E}}, \tag{4}$$

$$\bar{E} = -\frac{\partial \log Q}{\partial \beta}, \tag{5}$$

$$\frac{\partial^2 \log \Omega}{\partial \bar{E}^2} \leq 0. \tag{6}$$

Equations (3) and (4) establish that $\log \Omega(E, V, T)$ and $\log Q(\beta, V, N)$ are Legendre pairs; Equation (6) states that $\log \Omega$ is concave. In addition, any probability distribution p_i that could be assigned to microstate i under fixed $(\bar{E}, V, N)$ satisfies the inequality,

$$-\sum_i p_i \log p_i \leq \log \Omega, \tag{7}$$

with the equal sign only for the canonical distribution in Equation (1). This inequality is the statistical expression of the second law. If we identify $k_B \log \Omega$ with entropy and $-(\log Q)/\beta$ with free energy Equations (3)–(6) represent the familiar relationships of classical thermodynamics. Along with Equations (2) and (7), which provide the probabilistic context, the above set comprises the core relationships of statistical thermodynamics. The physical assumptions and postulates that produce these results can be found in any standard textbook (for example [3]). We will now show that this network of mathematical relationships arises naturally via a probabilistic construction that makes no reference to physics and endows any probability distribution $f(x)$, $x \geq 0$ with the thermodynamic relationships shown here.

3. Theory

3.1. Random Sampling

Consider the continuous probability distribution $h_0(x) \geq 0$, $x \in (x_a, x_b)$, normalized to unit area. We define a discrete grid $x_i = x_a + (i-1)\Delta$ with $\Delta = (x_b - x_a)/K$, $i = 1, 2 \cdots K+1$, such that the probability to sample a value of x in the ith interval is

$$p_i = h_0(x_i)\Delta, \tag{8}$$

if Δ is sufficiently small. We sample N values from h_0 and construct the frequency distribution $\mathbf{n} = (n_1, n_2, \cdots)$, where n_i is the number of sampled values that lie in the ith interval. The probability to observe distribution $\mathbf{n}$ in a random sample of size N is

$$P(\mathbf{n}|\mathbf{p}, N) = N! \prod_i \frac{p_i^{n_i}}{n_i!}, \tag{9}$$

and its logarithm is

$$\log P(\mathbf{n}|\mathbf{p}, N) = -\sum_i n_i \log \frac{n_i}{p_i N} + O(\log N), \tag{10}$$

where $\mathbf{p} = (p_1, p_2 \cdots)$. We define $h(x_i) = n_i/N\Delta$ and take the limit $\Delta \to 0$, $N \to \infty$ in Equation (10). We then have $P(\mathbf{n}|\mathbf{p}, N) \to \delta P(h|h_0, N)$ and

$$\frac{\log \delta P(h|h_0, N)}{N} = -\int h(x) \log \frac{h(x)}{h_0(x)} dx \doteq -D(h||h_0), \tag{11}$$

where $\delta P(h|h_0, N)$ is the probability to sample region $(h, h + \delta h)$ in the continuous space of distributions, while taking a random sample of size N from h_0 (all integrals are understood to be taken over the domain of h_0). Any probability distribution $h(x)$ defined in the domain of h_0 may materialize in a random sample taken from h_0. Clearly, the most probable distribution in this space is h_0 and indeed h_0 maximizes Equation (11). For all other distributions we must have $\delta P(h|h_0, N) \leq \delta P(h_0|h_0, N) = 1$, or

$$D(h||h_0) \geq 0, \tag{12}$$

with the equal sign only for $h = h_0$. The probability in the limit $N \to \infty$ to obtain h_0 relative to the probability to obtain any other distribution is

$$\frac{\delta P(h_0|h_0, N)}{\delta P(h|h_0, N)} = e^{ND(h||h_0)} \to \infty. \tag{13}$$

Accordingly, h_0 is overwhelmingly more probable than any other distribution in its domain.

These results make contact with a broader mathematical literature. The quantity $D(h||h_0)$ in Equation (11) is the relative entropy (Kullback-Leibler divergence) of distribution h with respect to h_0, and plays an important role in Information Theory [13–15]; Equation (12) is the Gibbs inequality, a well known property of relative entropy; the relationship between relative entropy and the probability of a sample drawn from h_0 is a known result in the theory of large deviations [16]. The key point we take from these results is that the process of sampling distribution h_0 establishes a probability space of distributions with the same domain as h_0—these are the distributions obtained as samples. The Gibbs inequality states the elementary fact that the most probable distribution in this space is h_0. We will now generalize this probability space and the Gibbs inequality.

3.2. Biased Sampling

Random sampling always converges to the distribution from which the sample is taken; the probability of all other distributions vanishes as $N \to \infty$. We now modify the sampling process in order to obtain some different limiting distribution h^* while still sampling from h_0. We do this by applying a bias, such that a random sample of size N from h_0 is accepted with probability proportional to $W[Nh]$, where Nh is the frequency distribution of the sample and W is a functional with the homogeneous property $\log W[Nh] = N \log W[h]$. We require homogeneity so that the limiting distribution is independent of N when $N \to \infty$. By virtue of homogeneity, $\log W$ is written as

$$\log W[h] = \int h(x) \log w(x;h) dx, \tag{14}$$

where $\log w(x;h)$ is the variational derivative of $\log W[h]$ with respect to h. The probability to obtain a sample with distribution h under this biased sampling is

$$P(h|\mathbf{p}, W, N) = \frac{W[Nh_i]}{r^N} \left(N! \prod_i \frac{p_i^{n_1}}{n_i!} \right), \tag{15}$$

where r^N is a normalizing constant; the logarithm of this probability in the continuous limit is

$$\frac{\log \delta P(h|h_0, W, N)}{N} = -\int h(x) \log \frac{h(x)}{w(x;h) h_0(x)} dx - \log r. \tag{16}$$

We define the probability functional

$$\log \varrho[h|h_0, W] \doteq -\int h(x) \log \frac{h(x)}{w(x;h) h_0(x)} dx - \log r, \tag{17}$$

so that the probability to observe a distribution within $(h, h + \delta h)$ in a biased sample taken from h_0 is $\delta P(h|h_0, N) = \varrho^N[h|h_0, W]$. The ratio of the probability to sample the most probable distribution h^* relative to that for any other distribution in the continuous limit is

$$\frac{\delta P(h^*|h_0, W, N)}{\delta P(h|h_0, W, N)} = \left(\frac{\varrho[h^*|h_0, W]}{\varrho[h|h_0, W]} \right)^N \to \infty. \tag{18}$$

As in random sampling, the most probable distribution is overwhelmingly more probable than any other feasible distribution. Then we must have

$$\varrho[h|h_0, W] \leq 1, \tag{19}$$

with the equal sign only for the most probable distribution h^*. This distribution is (see Supplementary Information).

$$h^*(x) = w(x;h^*) \frac{h_0(x)}{r}, \tag{20}$$

with r determined by normalization. If we choose $w(x;h) = f(x)/h_0(x)$, where f is any other normalized distribution in the domain of h_0, we obtain $h^* = f$. Therefore, a suitable bias can always be constructed such that *any* distribution in the domain of h_0 may be obtained as the most probable distribution by biased sampling of h_0; conversely, any distribution h_0 may be used to generate a sample of any other distribution f over the same domain by biased sampling.

3.3. Canonical Sampling

We now choose the generating distribution h_0 to be the normalized exponential distribution with parameter β,

$$h_0(x) = \beta e^{-\beta x}; \quad 0 \leq x < \infty, \tag{21}$$

and write the probability functional ϱ in Equation (17) as

$$\log \varrho[h|W, \beta] = -\int h(x) \log \frac{h(x)}{w(x;h)} dx - \beta \bar{x} - \log q, \tag{22}$$

where $q = r/\beta$ and $\bar{x}$ is the mean of $h(x)$. We call this probability space *canonical*. The probability of h is $\varrho^N[h|W, \beta]$ and by the same argument that led to Equation (19) we now have

$$\varrho[h|W, \beta] \leq 1. \tag{23}$$

The equal sign defines the most probable distribution h^*; this distribution is

$$h^*(x) = w(x; h^*) \frac{e^{-\beta x}}{q}. \tag{24}$$

The parameter q is fixed by the normalization condition and satisfies

$$\bar{x} = -\frac{d \log q}{d\beta}. \tag{25}$$

(Details are given in Supplementary Information).

3.4. Microcanonical Sampling

Next we define the *microcanonical* space as the subset of distributions with fixed mean $\bar{x}$. The generating distribution is again the exponential function, which we now write as

$$h_0(x) = \frac{e^{-x/\bar{x}}}{\bar{x}}, \tag{26}$$

with $\bar{x}$ fixed. The probability to observe distribution h while sampling h_0 is still given by Equation (16) except that r is replaced with a new normalizing factor r'. We define the microcanonical probability functional

$$\log \varrho[h|W, \bar{x}] = -\int h(x) \log \frac{h(x)}{w(x;h)} dx - \log \omega, \tag{27}$$

with $\log \omega = 1 + \log \bar{x} + \log r'$ and write the probability of h as $\varrho^N[h|W; \bar{x}]$. The argument that produced Equations (19) and (23) now gives

$$\varrho[h|W, \bar{x}] \leq 1. \tag{28}$$

This functional is maximized by the same distribution h^* that maximizes the canonical functional, Equation (24), except that both q and β are now Lagrange multipliers and are fixed by normalization and by the known mean $\bar{x}$. As in the canonical case, h^* is overwhelmingly more probable than any other distribution in the microcanonical space and its mean satisfies Equation (25). We insert Equation (24) into (28) to obtain

$$\log \omega = S[h^*] + \log W[h^*], \tag{29}$$

where $S[h^*]$ is the Gibbs–Shannon entropy of the most probable distribution,

$$S[h^*] = -\int_0^\infty h^*(x) \log h^*(x) dx. \tag{30}$$

Substituting Equation (24) for h^* in (29) we obtain a relationship between ω, β, q and $\bar{x}$:

$$\log \omega = \beta \bar{x} + \log q. \tag{31}$$

In combination with Equation (25), this result defines $\log \omega(\bar{x})$ as the Legendre transformation of $q(\beta)$ with respect to β. By the reciprocal property of the transformation we then have

$$\beta = \frac{d \log \omega}{d\bar{x}}. \tag{32}$$

Given Equation (31), the canonical probability functional in Equation (22) and the microcanonical functional in Equation (27) are seen to be the same. The difference is that in canonical maximization $\bar{x}$ is a floating parameter, whereas in the microcanonical maximization it is held constant. Both functionals are maximized by the same distribution and have the same β, q, ω at same $\bar{x}$: the two ensembles are equivalent. Finally, the maximization of the microcanonical functional implies that $\varrho[h; W, \bar{x}]$ is a concave functional in h. It follows that $\log \omega$ is a concave function of $\bar{x}$, therefore we must have

$$\frac{d^2 \log \omega}{d\bar{x}^2} = \frac{d\beta}{d\bar{x}} \leq 0. \tag{33}$$

The details are shown in Supplementary Information.

4. Generalized Statistical Thermodynamics

These results can be summarized in the form of the following theorem:

Theorem 1. *Given normalized distribution $f(x)$, $x \geq 0$, with mean $\bar{x}$, it is possible to construct a functional W such that:*

(a) All distributions $h(x)$, $x \geq 0$, with mean $\bar{x}$ satisfy the inequality

$$\log W[h] - \int_0^\infty h(x) \log h(x) dx \leq \log \omega \tag{34}$$

with the equal sign only for $h = f$, a condition that defines ω;

(b) f can be expressed in canonical form as

$$f(x) = w(x) \frac{e^{-\beta x}}{q}, \tag{35}$$

where $\log w$ *is the variational derivative of* $\log W[f]$*; and*

(c) parameters $\bar{x}$, β, q and ω satisfy

$$\bar{x} = -\frac{d \log q}{d\beta}, \tag{36}$$

$$\beta = \frac{d \log \omega}{d\bar{x}}, \tag{37}$$

$$\log \omega = \beta \bar{x} + \log q, \tag{38}$$

$$\frac{d^2 \log \omega}{d\bar{x}^2} \leq 0. \tag{39}$$

The existence of W is established by the fact that the functional

$$\log W[h] = \int_0^\infty h(x) \log f(x) dx, \tag{40}$$

satisfies the theorem. This is a linear functional whose derivative is $\log f$ for all h. More generally, any homogeneous functional $\log W[h]$ of degree 1, linear or non-linear, whose derivative at $h = f$ is given by

$$\left.\frac{\delta \log W[h]}{\delta h}\right|_{h=f} = \log f(x) + a_0 + a_1 x \doteq \log w(x), \tag{41}$$

where a_0 and a_1 satisfy

$$\frac{da_0}{da_1} = -\bar{x}, \tag{42}$$

but are otherwise arbitrary, also satisfies the theorem. The inequality in Equation (39) follows from the concave requirement that ensures the maximization of Equation (34).

We recognize Equation (35) as the canonical distribution of statistical mechanics, Equations (36)–(38) and (33), which relate its parameters, as the core set of thermodynamic relationships, and Equation (34) as the inequality of the second law. The probabilistic interpretation is that any distribution f may be obtained as the most probable distribution under a probability measure defined via a suitable functional W. Whereas in statistical mechanics the central stochastic variable is the mechanical microstate, in generalized thermodynamics it is the probability distribution itself. Thermodynamics may be condensed into the microcanonical inequality in Equation (34), a generalized expression of the second law that defines the most probable distribution in the microcanonical space. All relationships between ω (microcanonical partition function), q (canonical partition function), β (generalized inverse temperature) and $\bar{x}$ follow from the maximization of this inequality and have equivalents in familiar thermodynamics. The derivatives $d \log q / d\beta$ and $d \log \omega / d\bar{x}$ in Equations (36) and (37) may be viewed as equations of change along a path ("process") in the space of distributions under fixed bias W. This path is described parametrically in terms of $\bar{x}$ and represents a nonstationary stochastic process. We call this process *quasistatic*—a continuous path of distributions that maximize locally the thermodynamic functional.

4.1. Contact with Statistical Mechanics

The obvious way to make contact with statistical mechanics is to take f to be the probability of microstate at fixed temperature, volume and number of particles. The postulate of equal a priori probabilities fixes the selection functional and its derivative to $W = w = 1$; if we identify x as the energy E_i of microstate i, β as $1/k_B T$, q as the thermodynamic canonical partition function, ω as the thermodynamic microcanonical partition function, Equations (24)–(33) map to standard thermodynamic relationships. From Equation (29) we obtain $\varrho = e^{S[h]}/\omega$: the canonical probability f maximizes entropy and thus we obtain the second law.

This is not the only way to establish contact with statistical mechanics. We may choose f to be some other probability distribution, for example, the probability to find a *macroscopic* system of fixed (T, V, N) at energy E. We write the energy distribution in the form of Equation (24) with w, β and q to be determined. From Equations (25), (31) and (32) with $\bar{x} = \bar{E}$ we make the identifications $\beta \to 1/k_B T$, $\log q \to -F/k_B T$ (free energy), $\log \omega \to$ thermodynamic entropy. To identify w we require input from physics and this comes via the observation that the probability density of macroscopic energy E is asymptotically a Dirac delta function at $E = \bar{E}$. Then $S[f] = 0$ (this is the entropy of the energy distribution, not to be confused with thermodynamic entropy). From Equations (14) and (29) we find $\log W[f] = \log w(x; f) = \log \omega$, and conclude that $\log \omega$ is the thermodynamic entropy. This establishes correspondence between generalized thermodynamics and macroscopic (classical) thermodynamics. If we further postulate, again motivated by physics, that $w(E)$ is the number of microstates under fixed volume and number of particles, we establish the microscopic connection. Since $f(E)$ is proportional to the number of microstates with energy E and individual microstates are unobservable, we may as well ascribe equal probability to all microstates. Thus we recover the postulate of equal a priori probabilities (statistical thermodynamics). Finally, by adopting a physical model of microstate, classical, quantum or other, we obtain classical statistical mechanics, quantum

statistical mechanics or yet-to-be-discovered statistical mechanics, depending on the model. In all cases the thermodynamic calculus is the same, only the enumeration of microstates—that is, W— depends on the physical model.

4.2. What is W?

Once the selection functional W is specified the most probable distribution is fixed and all canonical variables become known functions of $\bar{x}$. But what *is* W? The selection functional is a placeholder for our knowledge, hypotheses and model assumptions about the stochastic processes that gives rise to the probability distribution of interest. This knowledge fully specifies the distribution. The opposite is not true: given distribution f there is an infinite number of functionals W that produce that distribution as the most probable distribution in their probability space. This nonuniquness is a feature, not a bug: it allows models that are quite different in their details to produce the same final distribution. Here is an example. The unbiased functional $W[h] = w(x) = 1$ produces the exponential distribution

$$h^*(x) = \frac{e^{-\beta x}}{q}, \tag{43}$$

with canonical parameters

$$\beta = 1/\bar{x}, \quad q = \bar{x}, \quad \log \omega = 1 + \log \bar{x}. \tag{44}$$

Now consider the nonlinear selection functional

$$\log W[h] = S[h] = -\int_0^\infty h(x) \log h(x) dx, \tag{45}$$

whose logarithm is equal to entropy. The corresponding microcanonical probability functional is obtained by inserting this into Equation (27),

$$\log \varrho[h|W, \bar{x}] = -2 \int_0^\infty h(x) \log h(x) dx - \log \omega \tag{46}$$

and is maximized by (see Supplementary Information)

$$h^*(x) = w(x) \frac{e^{-\beta x}}{q}, \tag{47}$$

with

$$w(x) = \bar{x} e^{x/\bar{x}}, \ \beta = 2/\bar{x}, \ q = \bar{x}^2, \ \log \omega = 2 + 2 \log \bar{x}. \tag{48}$$

We have arrived at the *exponential* distribution, the same distribution that is obtained by the unbiased functional $w(x) = 1$, but with different canonical parameters because the probability space from which it arises is now different. If all we know is that the probability distribution in a stochastic process is exponential, it is not possible to determine whether it was obtained using $W[h] = 1$, $W[h] = e^{S[h]}$, or any other functionals that is capable of reproducing the exponential distribution. While the selection bias identifies the most probable distribution uniquely, the opposite is not true.

The selection functional represents external input to thermodynamics and is fixed by the rules that govern the stochastic process that produces the distribution in question. In the case of statistical mechanics it is fixed by the postulate of equal a priori probabilities. In another example, recently given for stochastic binary clustering, it is fixed by the aggregation kernel, a function that determines the aggregation probability between clusters of different sizes [17]. The selection functional is the contact point between generalized statistical thermodynamics—a mathematical theory for generic distributions—and *physics*, i.e., our knowledge in the form of model assumptions and postulates about the process that gives rise to the observed distribution. It is interesting to point out that the variational derivative w in Equation (27) appears in the form of Bayesian prior [18]. In the context of generalized

thermodynamics w is not a prior distribution—although it might if $a_0 = a_1 = 0$ in Equation (41). In general, w is a non normalizable derivative of the functional that represents our knowledge about the process, an improper prior that points nonetheless to a proper distribution.

5. Thermodynamic Sampling of Distributions

We have shown that any distribution $f(x)$ defined in $\mathbb{R}_+$ can be viewed as the most probable distribution in an appropriately constructed probability space. Here we will show that any distribution f in this domain can be obtained as the equilibrium distribution of reacting clusters under an appropriately constructed equilibrium constant. Consider a population of M identical particles ("monomers") distributed into N clusters and let $\mathbf{m} = (m_1, m_2 \cdots, m_N)$ be an ordered list of N cluster masses with total mass M such that m_k is the mass of the kth cluster in the list ("configuration"). The complete set of configurations with N clusters and total mass M comprises the cluster ensemble (M, N). Let $\mathbf{n} = (n_1, n_2 \cdots)$ be the size distribution of the clusters in configuration $\mathbf{m}$ such that n_i is the number of clusters with i monomers. With $M, N \to \infty$ at fixed $M/N = \bar{x}$, the cluster ensemble contains every discrete distribution $h_i = n_i/N$ with mean $\bar{x}$. We now construct the following stochastic process: given a configuration $\mathbf{m}$, pick two clusters at random, merge them, then split them back into two clusters at random. This amounts to an exchange of mass between two clusters that is represented schematically by the reaction

$$m_i + m_j \to m_i' + m_j' \tag{49}$$

and transforms the parent configuration $\mathbf{m}$ into an new configuration $\mathbf{m}'$ with the same number of clusters N and total mass M. This process may also be represented as a reaction that transforms a parent configuration into an offspring,

$$\mathbf{m} \xrightarrow{K} \mathbf{m}'. \tag{50}$$

We define the equilibrium constant of this reaction as

$$K_{\mathbf{m}\to\mathbf{m}'} = \frac{W(\mathbf{n}')}{W(\mathbf{n})}, \tag{51}$$

where $\mathbf{n}'$ and $\mathbf{n}$ are the cluster size distributions of the product and reactant configurations, respectively, and $W(\mathbf{n})$ is the selection functional applied to distribution $\mathbf{n}$. The change $\delta\mathbf{n}$ of the corresponding distributions upon the exchange reaction is a change of -1 in the number of cluster masses m_i and m_j on the reactant side, and $+1$ for cluster masses on the product side. By virtue of the homogeneous property of $\log W$, its change for large M and N is a differential that can be expressed in terms of the derivatives of $\log W$

$$\log W(\mathbf{n}') - \log W(\mathbf{n}) = -\log w(m_i) - \log w(m_j) + \log w(m_i') + \log w(m_j'), \tag{52}$$

where $\log w$ is the functional derivative of $\log W$ evaluated in distribution $\mathbf{n}$. Using this result the equilibrium constant becomes

$$K_{\mathbf{m}\to\mathbf{m}'} = \frac{w(m_i')w(m_j')}{w(m_i)w(m_j)}. \tag{53}$$

This has the standard form of an equilibrium constant for the reaction in Equation (49). We may identify $w(x)$ as the "fugacity" of species x and "species" as a cluster with mass x. The reaction can be simulated by Monte Carlo using the Metropolis transition probabilities

$$P_{\mathbf{n}\to\mathbf{n}'} = \begin{cases} \texttt{rnd} & \text{if } \texttt{rnd} \le K_{\mathbf{n}\to\mathbf{n}'}, \\ 1 & \text{if } \texttt{rnd} > K_{\mathbf{n}\to\mathbf{n}'}, \end{cases} \tag{54}$$

where `rnd` is a uniform random number in $(0, 1)$. This forms a reducible Markov process that samples the microcanonical space of distribution $\mathbf{n}$ with fixed zeroth order moment N and first moment M. Its stationary distribution is [19]

$$h^*(x) = w(x)\frac{e^{-\beta x}}{q}. \tag{55}$$

where $\log w(x)$ is the functional derivative of $\log W$ evaluated at $h = h^*$ and the parameters β and q are obtained by solving the set of equations

$$q = \int_0^\infty w(x)e^{-\beta x}dx, \tag{56}$$

$$\bar{x} = \frac{1}{q}\int_0^\infty xw(x)e^{-\beta x}dx. \tag{57}$$

With $W[h] = w[x] = 1$ we obtain the exponential distribution, which implies that the exchange reaction with equilibrium constant $K = 1$ for all transitions is equivalent to unbiased sampling from an exponential distribution with fixed mean $\bar{x} = M/N$.

Once the selection functional W is given the most probable distribution is fixed and may be obtained either by simulation or in many cases analytically. We will now construct W such that the most probable distribution is any distribution f defined in $\mathbb{R}_+$. We construct the linearized selection functional

$$\log W[h] = \int_0^\infty h(x)\log w(x)dx \tag{58}$$

with w from Equation (41), which we write in the form

$$w(x) = f(x)e^{a_0 + a_1 x} \tag{59}$$

and a_0 and a_1 arbitrary constants. It is easy to show that the selection of a_0 and a_1 is immaterial because both constants drop out of Equation (53). If we choose $a_0 = a_1$, then $w(x) = f(x)$; alternatively, we may choose these constants so as to obtain simpler forms for $w(x)$. We demonstrate the construction of w with three examples using the exponential, the Weibull, and the uniform distribution.

1. Exponential distribution

$$f(x) = e^{-x/\bar{x}}/\bar{x}. \tag{60}$$

The function w is

$$w(x) = \frac{e^{-x/\bar{x} + a_0 + a_1}}{\bar{x}}. \tag{61}$$

Choosing $a_0 = \log \bar{x}$, $a_1 = 1/\bar{x}$ we obtain $w_{\text{exp}}(x) = 1$, which represents the unbiased selection functional.

2. Weibull distribution

$$f(x) = \left(\frac{k}{\lambda}\right)\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}.$$

Using $a_0 = k\log\lambda - \log k$ and $a_1 = 0$ in Equation (59) we obtain

$$w_{\text{Weibull}}(x) = x^{k-1}e^{-(x/\lambda)^k}. \tag{62}$$

3. Uniform distribution

$$f(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise.} \end{cases} \tag{63}$$

With $a_0 = a_1 = 0$ we obtain

$$w_{\text{uniform}}(x) = f(x). \tag{64}$$

We implement thermodynamic sampling using Monte Carlo. We begin with an ordered list of N integers $i > 0$ whose sum is M. We then pick two numbers at random and implement a random exchange reaction to produce a new pair of integers with the same combined sum. The new pair replaces the old with acceptance probability computed according to Equation (54) using K_{eq} from Equation (53) and the function $w(x)$ obtained above. Following a successful trial we calculate the distribution of the current configuration. The mean distribution is obtained by averaging over a large number of trials. For these simulations $N = 100$, $M = 3000$, $\bar{x} = 30$, and the mean distribution is calculated over 20,000 trials. As we discuss elsewhere, the mean distribution and the most probable distribution converge to each other unless the system exhibits phase separation [17,19,20]. The results in Figure 1 make it clear that thermodynamic sampling converges indeed to the distribution for which the w function was derived. Any discrete distribution h_i, and with proper scaling, any continuous distribution $h(x)$, may be associated with the equilibrium distribution of reacting clusters under a suitable equilibrium constant.

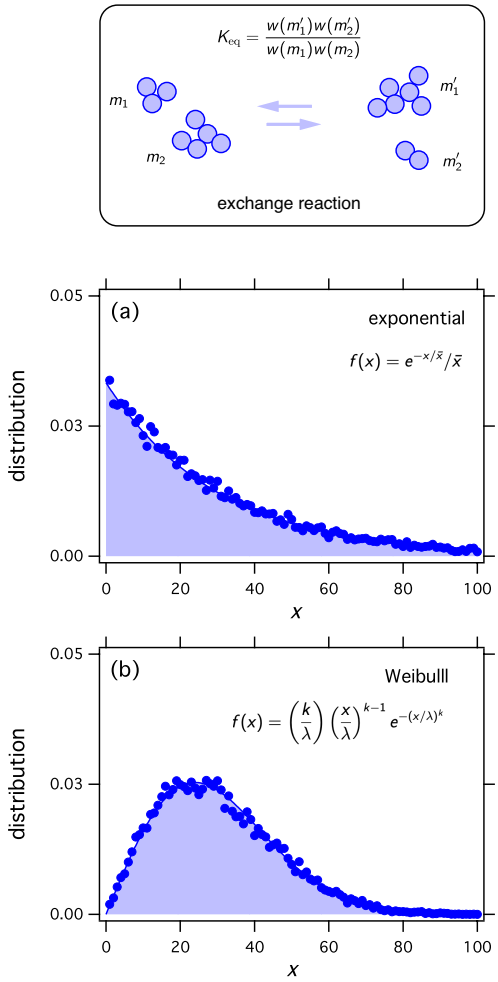

Figure 1. *Cont.*

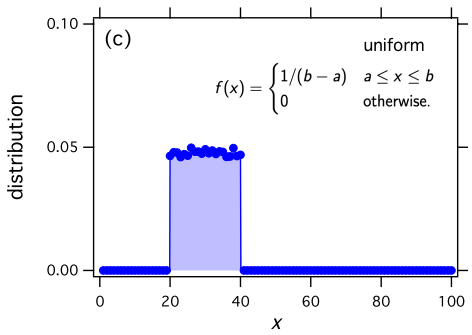

Figure 1. The exchange reaction transfers mass between two clusters and samples the space of all distributions with fixed number of clusters N and fixed total number of monomers M. We may construct the equilibrium constant of this reaction so as to to obtain any desired equilibrium distribution. Any distribution $f(x)$, $x \geq 0$, may be obtained as the equilibrium distribution. In this example we obtain (**a**) the exponential distribution; (**b**) the Weibull distribution with $\lambda = 33.8514$, $k = 2$; and (**c**) the uniform distribution between $a = 20$ and $b = 40$. In all cases $\bar{x} = 30$.

The selection functionals constructed by the procedure discussed here apply the variational derivative at f to all distributions h, i.e., they are linearized at the most probable distribution. Any nonlinear functional $\log W$ with the same derivative at $h = f$ will produce the same distribution as the stationary distribution under exchange reactions. One example is the entropic functional in Equation (45), a nonlinear functional that produces the exponential distribution. Even though the entropic and unbiased functionals both produce the same distribution (Figure 2a), their corresponding ensembles are distinctly different because each functional assigns different probabilities to the distributions of the ensemble. This difference can be seen in the fluctuations (Figure 2b). The entropic functional is more selective than the unbiased, which picks configurations with equal probability. Accordingly, fluctuations in the entropic ensemble have narrower distribution. This can be clearly seen in Figure 2b that shows the fluctuations in the number of monomers for the entropic and the unbiased functionals.

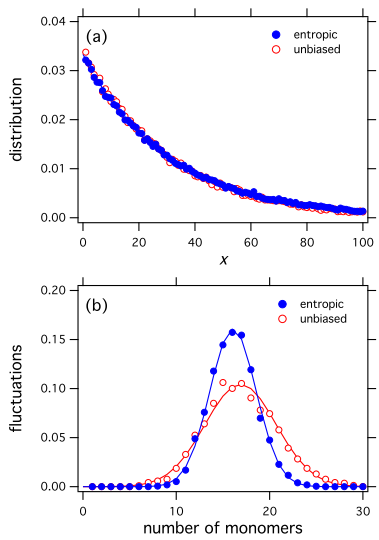

Figure 2. (**a**) The entropic selection functional, $W[h] = e^{S[h]}$, and the unbiased functional, $W[h] = 1$, both produce the same equilibrium distribution (exponential). Nonetheless the two selection functionals represent distinctly different ensembles, as can be seen in fluctuations of the number of monomers (**b**). The entropic functional is more selective than the unbiased and produces a tighter distribution of fluctuations.

6. Conclusions

Stripped to its core, what we call statistical thermodynamics is a mapping between a probability distribution f and a set of functions, $\{w, \beta, q, \omega\}$ from which the distribution may be reconstructed. What we call classical thermodynamics is the set of relationships among $\{\beta, q, \omega, \bar{x}\}$—relationships that are the same for all distributions. What we call second law is the variational condition that identifies the most probable distribution in the domain of feasible distributions. What we call quasistatic process is a path in the space of distributions under fixed W. Physics enters through W. This generic mathematical formalism applies to any distribution. To use an analogy, thermodynamics is a universal grammar that becomes a language when applied to specific problems. It is a fitting coincidence—or perhaps an inevitable consequence—that it was the human desire to maximize the amount of useful work in the steam engine that would eventually make contact with the variational foundation of thermodynamics. Gibbs's breakthrough was to connect thermodynamics to a probability distribution, and that of Shannon and Jaynes to transplant it outside physics. In the time since, the vocabulary of statistical thermodynamics has felt intuitively familiar across disciplines in a déjà vu sort of manner, even as its grammar remained undeciphered. This intuition can now be understood: The common thread that runs through every discipline that has adopted the thermodynamic language is an underlying stochastic process, and where there is probability, there is statistical thermodynamics.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflict of interest.

References

1. *The Concise Oxford English Dictionary*, 11th ed.; Oxford University Press: Oxford, United Kingdom. 2008.
2. Callen, H. *Thermodynamics and Introduction to Thermostatistics*, 2nd ed.; Wiley: Hoboken, NJ, USA, 1985.
3. Hill, T.L. *Statistical Mechanics Principles and Selected Applications*; Reprint of the 1956 Edition; Dover: Mineola, NY, USA, 1987.
4. Huang, K. *Statistical Mechanics*, 2nd ed.; Wiley: New York, NY, USA, 1963.
5. Gibbs, J.W. *Elementary Principles in Statistical Mechanics*; Reprint of the 1902 Edition; Ox Bow Press: Woodbridge, CT, USA, 1981.
6. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
7. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. doi:10.1103/PhysRev.106.620. [CrossRef]
8. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97, doi:10.1103/RevModPhys.74.47. [CrossRef]
9. Harte, J.; Zillio, T.; Conlisk, E.; Smith, A.B. Maximum Entropy and the State-Variable approach to macroecology. *Ecology* **2008**, *89*, 2700–2711, doi:10.1890/07-1369.1. [CrossRef] [PubMed]
10. Durrett, R. Stochastic Spatial Models. *SIAM Rev.* **1999**, *41*, 677–718, doi:10.1137/S0036144599354707. [CrossRef]
11. Timme, N.M.; Lapish, C. A Tutorial for Information Theory in Neuroscience. *eNeuro* **2018**. [CrossRef] [PubMed]
12. Voit, J. *The Statistical Mechanics of Financial Markets*; Springer: Berlin/Heidelberg, Germany, 2005; doi:10.1007/b137351.
13. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Statist.* **1951**, *22*, 79–86, doi:10.1214/aoms/1177729694. [CrossRef]
14. Gray, R.M. *Entropy and Information Theory*, 2nd ed.; Springer: New York, NY, USA, 2011.
15. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley Interscience: Hoboken, NJ, USA, 2006.
16. Touchette, H. The large deviation approach to statistical mechanics. *Phys. Rep.* **2009**, *478*, 1–69. doi:10.1016/j.physrep.2009.05.002. [CrossRef]
17. Matsoukas, T. Statistical Thermodynamics of Irreversible Aggregation: The Sol-Gel Transition. *Sci. Rep.* **2015**, *5*, 8855. [CrossRef] [PubMed]

18. Jaynes, E. Prior Probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241, doi:10.1109/TSSC.1968.300117. [CrossRef]
19. Matsoukas, T. Statistical thermodynamics of clustered populations. *Phys. Rev. E* **2014**, *90*, 022113, doi:10.1103/PhysRevE.90.022113. [CrossRef] [PubMed]
20. Matsoukas, T. Abrupt percolation in small equilibrated networks. *Phys. Rev. E* **2015**, *91*, 052105. doi:10.1103/PhysRevE.91.052105. [CrossRef] [PubMed]

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).