Article

Internal Energy, Fundamental Thermodynamic Relation, and Gibbs’ Ensemble Theory as Emergent Laws of Statistical Counting

Hong Qian

Department of Applied Mathematics, University of Washington, Seattle, WA 98195-3925, USA; hqian@uw.edu

Abstract: Statistical counting ad infinitum is the holographic observable to a statistical dynamics with finite states under independent and identically distributed $N$ sampling. Entropy provides the infinitesimal probability for an observed empirical frequency $\hat{\nu}$ with respect to a probability prior $p$, when $\hat{\nu} \neq p$ as $N \to \infty$. Following Callen’s postulate and through Legendre–Fenchel transform, without help from mechanics, we show that an internal energy $u$ emerges; it provides a linear representation of real-valued observables with full or partial information. Gibbs’ fundamental thermodynamic relation and theory of ensembles follow mathematically. $u$ is to $\hat{\nu}$ what chemical potential $\mu$ is to particle number $N$ in Gibbs’ chemical thermodynamics, what $\beta = T^{-1}$ is to internal energy $U$ in classical thermodynamics, and what $\omega$ is to $t$ in Fourier analysis.

Keywords: emergent phenomenon; entropy; information; internal energy; probability theory; statistic; thermodynamics

1. Introduction

It is a pleasure to be a part of this celebration for Signe Kjelstrup. She has made significant contributions to nonequilibrium thermodynamics, in both theory and applications that include electrochemistry, transport in heterogeneous media, and T. L. Hill’s small systems [1–3]. In this work we extend Gibbs’ and Hill’s approach to equilibrium thermodynamics [4] and show that the new logical path via the “crucial step” advocated in [5] is in fact a consequence of a limit theorem [6] in the mathematical theory of probability [7,8]. The results perfectly fit P. W. Anderson’s notion of emergent phenomenon [9].

Sometimes a mathematical transform can provide a fundamental concept beyond just being a technique for solving a problem, and through which a new representation of a natural phenomenon emerges. A case in point is the Fourier transform (FT) that leads to the theory of harmonics in music instruments [10] and the very concept of optical spectrum. FT represents a function of time $f(t)$ in terms of $\tilde{f}(\omega)$, where $\omega$ is introduced as a novel notion, the temporal frequency of a sinusoidal oscillatory component in time [11]. The solutions to a large class of problems in differential calculus involving $t$ can be very efficiently expressed through FT.

We show in the present paper that the fundamental notion of internal energy first appeared in the theory of thermodynamics in the 19th century, collectively developed by J. R. von Mayer, W. Rankine, R. Clausius, and W. Thomson among many others [12], is a concept that can be understood, and generalized, in statistical counting. The transformation in question is the Legendre–Fenchel transform (LFT) [13,14], a more refined mathematical formulation of the traditional Legendre transform [15].

When a simple statistical analysis is carried out on a set of data, correlated or not, it is usually supposed that they are from an identical probability distribution. One of the best understood systems that exhibit an invariant probability is the ergodic dynamical system [16]. The ergodic theory of classical Hamiltonian dynamics has been an intense research area in both physics and mathematics for more than a century [17,18]. Even
when the data are from seemingly different “objects”, say different individuals within a biological species, it is understood that an ergodic mating or mutational process is behind the statistical practice; and the conclusions drawn are most meaningful in this regard. Such an ergodic stochastic dynamic perspective has transformed cell biology through the notion of phenotypic switching in recent years [19].

2. Energetic Theory of Statistical Counting

Let us consider the repeated statistical samples ad infinitum of a system with finite state space \( S = \{0, 1, \cdots, n\} \). In the present work we shall restrict our discussion for independent and identically distributed (i.i.d.) samples. More general sampling of Markov data will be published elsewhere. The number counting \( \nu = (\nu_0, \cdots, \nu_n) \) with \( \nu_0 + \cdots + \nu_n = N \) and counting frequency \( \dot{\nu} = \nu/N \), not to be confused with the \( \omega \) in FT above, has a homogeneous degree 1 neg-entropy function with respect to a given probability prior \( p = (p_0, \cdots, p_n) \) [8,20]:

\[
\Phi(\nu) = \sum_{i=0}^{n} \nu_i \ln \left( \frac{\nu_i}{\sum_{j=0}^{n} \nu_j} \right). \tag{1}
\]

The Appendix A provides the mathematical origin of the non-negative \( \Phi(\nu) \) as a result of statistical counting. In information theory, it is interpreted as the “surprise” in observing the \( \nu \) under the assumption \( p \) [21,22]. It is a double-edged sword which tells the rareness of \( \nu \) (or \( \dot{\nu} \)) with respect to \( p \) or erroneous model \( p \) with respect to empirical \( \nu \). The Kolmogorov probability \( p \) has two rather different roles in statistical inference and in statistical physics. In the former it has been identified as only half to the other “half of probability theory as it is needed in current applications—the principles for assigning probabilities by logical analysis of incomplete information—is not present at all in the Kolmogorov system” [20].

The application of modern probability to statistical physics involves the limit of sample ad infinitum represented by \( N \to \infty \) [6]. In this case, the probability \( p \) is for all the systems with the same state space \( S \); it is not meant to be realistic for any particular system. It simply provides a “metric” under which each and every particular system has its own representation in terms of its complete information, \( \nu \). The \( \Phi(\nu) \) is introduced to further gauge the differences among systems with different \( \nu ' \)'s on the same \( S \); it becomes a “theory of everything.” Because the \( N \to \infty \) limit, there are no uncertainties in \( \dot{\nu} \); it is a definitive characterization of an i.i.d. statistical distribution with state space \( S \).

Therefore, statistical inference is about the mathematical model of a particular system, and statistical physics is about the mathematical representation of all systems with the same \( S \) under the supposition of i.i.d. data ad infinitum. The entropy in (1) is an emergent characterization in the limit of \( N \to \infty \), with the starting point in terms of generative models [9]. It provides the relationship between \( \nu \) and \( p \) in the sampling process. It is an Eulerian degree 1 homogeneous function of \( \nu \): \( \Phi(\lambda \nu) = \lambda \Phi(\nu) \). This fits naturally to the fundamental thermodynamic postulate formulated by H. B. Callen [23]. The LFT of \( \Phi \) as a function of the normalized \( \dot{\nu} \) then yields [13,14,24]:

\[
\Psi(\mathbf{u}) = \inf_{\dot{\nu}} \left\{ \sum_{i=0}^{n} \nu_i u_i + \Phi(\dot{\nu}) \right\} = -\ln \sum_{i=0}^{n} p_i e^{-u_i}, \tag{2a}
\]

with corresponding optimal \( \dot{\nu}^*(\mathbf{u}) \)

\[
(\dot{\nu}^*_i) = \frac{p_i e^{-u_i}}{\sum_{i=0}^{n} p_i e^{-u_i}}, \text{ and } u_i = -\left( \frac{\partial \Phi(\dot{\nu}^*)}{\partial \dot{\nu}_i} \right). \tag{2b}
\]

Note that the second equation in (2b) is obtained when one uses calculus to solve the infimum in (2a); this recovers the traditional Legendre transform. Normalizing \( \nu \) to \( \dot{\nu} \) induces a gauge freedom in (2), an arbitrary additive constant to \( u_i \). In statistical thermodynamics, the conjugate variable \( u_k \) introduced in Equation (2) has been interpreted as the
internal energy of the state \( k \), in \( k_BT \) unit \([25]\); then \( \dot{\nu} \cdot \mathbf{u} \) is the mean internal energy of “the statistical system”.

In a real-world laboratory working on a particular system, the \( \nu \) tends to infinity as \( N \to \infty \) but \( \dot{\nu} \) converges to the intrinsic property of the statistical system. In statistical inference, the assumed \( \mathbf{p} \), as a prior, then is expected to be replaced by the observed, real, posterior \( \dot{\nu} \) according to conditional probability and/or Bayesian statistical logic \([24,25]\). This concludes the statistical investigation of the particular system with respect to the type of observations. The neg-entropy function in (1) actually provides a meta-statistical theory for all possible observed \( \dot{\nu} \), assessing their respective infinitesimal probability (rate) with respect to the prior \( \mathbf{p} \) (see Appendix A).

3. Maximization of Entropy \( \Phi \) Under Constraint by Empirical Mean Value

However, The complete counting for the entire state space \( S \) in terms of empirical frequencies \( \dot{\nu} \) is only a gedankenexperiment. The significance of Gibbs’ ensemble theory is in dealing with observations from a small set of real-valued observables \( g_1(\nu), g_2(\nu), \ldots, g_J(\nu) \), where \( i \in S \) but \( J \ll n \). These \( g \)’s are random variables on the state space \( S \). In fact, their empirical mean values are linear combinations of the \( \dot{\nu} \):

\[
x_j = \sum_{i=0}^{n} \dot{\nu}_i g_j(i). \tag{3}
\]

To fix mathematical notations, we append \( g_0(i) = 1 \) and \( x_0 = 1 \), which represent the fact that \( \dot{\nu} \) is always normalized, and denote \((n+1) \times (J+1)\) matrix \( \mathbf{G}_J \) with elements

\[
(\mathbf{G}_J)_{ij} = \begin{cases} 1 & j = 0, \\ g_j(i) & j = 1, \cdots, J. \end{cases} \tag{4}
\]

Equation (3) shows that if all the \( g \)’s are linearly independent and \( J = n \), then one can solve the normalized \( \dot{\nu} \) uniquely from each set of \( x \)’s: \( \dot{\nu} = x\mathbf{G}_n^{-1} \). We refer to such a set of observables as holographic with full information. In the following discussion, we shall always imagine the \((g_1, \cdots, g_J)\) as the first \( J \) component of a holographic observable \((g_0, g_1, \cdots, g_n)\). When \( J < n \), there is missing information \([20,22,25]\).

With a set of observed values \( x' = (x_1, \cdots, x_J) \) in hand where \( J < n \), the maximum entropy principle (MEP) from classical thermodynamics \([23]\) and the contraction principle from the mathematical theory of probability \([8]\) assert that the most probable \( \dot{\nu}^* \) that is consistent with the set of \( x' \) corresponds to minimum neg-entropy:

\[
\dot{\nu}^* = \arg \inf_{\dot{\nu}} \{ \Phi(\dot{\nu}) \mid \dot{\nu} \mathbf{G}_J = x' \}. \tag{5}
\]

The entire Gibbs’ ensemble theory arises in solving the mathematical problem posed in Equation (5) through LFT. See Appendix A for its origin.

Entropy functions for different observables are different. First, for invertible \( \mathbf{G}_n \), one has the entropy function for the holographic observable \( x = (1, x_1, \cdots, x_n) \):

\[
\Phi_x(x) \equiv \Phi(x\mathbf{G}_n^{-1}). \tag{6}
\]

This is simply a change in the independent variables from \( \nu \) to \( x \). Then in terms of this entropy function \( \Phi_x(\dot{\nu}) \) (5) becomes

\[
\varphi(x') = \inf_{\dot{\nu}} \{ \Phi(\dot{\nu}) \mid \dot{\nu} \mathbf{G}_j = x' \} \tag{7}
\]

\[
= \inf_{x_1, x_1', \cdots, x_n} \{ \Phi_x(x) \mid x_1 = x_1', \cdots, x_J = x_J' \}.
\]
Intimately related to the generating function of a probability distribution, the LFT provides a powerful mathematical transform of the entropy functions $\Phi(\nu), \Phi_x(x),$ and $\varphi(x')$ in terms of their conjugates in the energy representation: Parallel to the $\Psi(u)$ in (2) are, 

$$\Psi_y(y) = \inf_x \left\{ \sum_{i=1}^{n} x_i y_i + \Phi_x(x) \right\},$$

and projection:

$$\psi(y') = \inf_y \left\{ \sum_{i=1}^{l} y_i y_i' + \varphi(y') \right\}.$$ (8)

These $\psi$'s are now related through linear transformation:

$$\Psi_y(y) = \Psi(G_n y),$$

and projection:

$$\Psi_y(y) = \Psi(y_i', \ldots, y_j, 0, \ldots, 0) = \Psi(G_J y'),$$ (11a)

$$\psi(y') = \inf_y \left\{ \sum_{i=0}^{n} p_i \exp \left[ - \sum_{j=1}^{l} g_j(i) y_j \right] \right\}. \quad (11b)$$

And finally, since $\psi$ is convex, the inverse LFT yields

$$-\varphi(x') = \inf_{y'} \left\{ \sum_{i=1}^{l} x_i y_i' - \psi(y') \right\}$$

$$= \left\{ \begin{array}{ll}
-\varphi &= y' \cdot \nabla \psi(y') - \psi(y') \\
x' &= \nabla \psi(y')
\end{array} \right. \quad (12)$$

The optimization in (5) is completely “solved” in closed form, through LFT and its inverse, as a parametric function in terms of $y'$ given in (12).

The equation $-\varphi = y' \cdot \nabla \psi - \psi$ in (12) should be recognized as a generalization of the celebrated “entropy = mean internal energy − free energy”, where

$$\left( \nabla \psi \right)_k = \frac{\sum_{i=0}^{n} p_i \exp \sum_{j=1}^{l} g_j(i) y_j}{\sum_{i=0}^{n} p_i \exp \sum_{j=1}^{l} g_j(i) y_j}$$

is the mean value of $g_k$ following Equation (11b), whose conjugate variable is $y'_k$. The identification of $u = G_J y'$ in (11a) with the first law of thermodynamics as formulated by Gibbs seems natural.

The $y_{j+1} = \cdots = y_n = 0$ in (11a) has a very clear thermodynamic interpretation: Since the conjugate variable $y$ are the partial derivatives of the entropy function $\Phi_x$ with respect to $x$, finding $x$'s with maximum entropy in Equation (7) is simply setting the corresponding $y = 0$, e.g., letting the entropic force be zero. For each independent observable $g_j, y_j$ is its “custom-designed” conjugate force and $y_j \times dx_j$ contributes a term to the internal energy as the “thermodynamic work” associated with $g_j$: The internal energy $u$ is a highly flexible, adaptive representation of the $\hat{\nu}$. When $J = n, u = G_n y$ and Equation (10) provides a complete “detailing” of the internal energy in terms of a set of holographic observables. MEP is for missing information [20].

4. Gibbs Distribution and Linear Algebraic Representation

There is a geometric picture associated with the above “thermodynamic analysis”. As we have stated, counting frequency ad infinitum $\hat{\nu}$ is a fundamental, intrinsic property of
an ergodic dynamical system. The space of all possible frequency distributions \( \hat{v} \), with \( \hat{v}_0 + \cdots + \hat{v}_n = 1 \), is a \( n \)-dimensional hyper-plane in the positive quadrant of \( \mathbb{R}^{n+1} \), known as a probability simplex \( \mathcal{M}_n \). For a given set of observables \( (g_1, \cdots, g_j) \), the \( \mathcal{M}_n \) is foliated by \( \hat{v} \mathcal{G}_j = \mathbf{x}' \) with different \( \mathbf{x}' \). On each leave of the foliation, there is the most probable \( \nu^*(x') \), which is located at the tangent point between the \( (n-1) \)-dimensional level set of the \( \Phi(\hat{v}) \) function. At this point, \( \nabla_v \Phi(\hat{v}) = -\mathbf{u}(\hat{v}) \) is the normal vector to the \( x' \)-leave in \( \mathbb{R}^{n+1} \), and \( \nabla \psi(x') = -\mathbf{y}' \) is its projection onto the \( j \)-manifold of \( x' \):

\[
\nu^*(x') = \begin{cases} 
\nu'_i t = \frac{1}{Z(y')} p_i \exp \left[- \sum_{j=1}^J g_j(i) y'_j \right] \\
x'_j = \frac{1}{Z(y')} \sum_{i=0}^n g_j(i) p_i \exp \left[- \sum_{j=1}^J g_j(i) y'_j \right] 
\end{cases} \tag{14a}
\]

in which

\[
Z(y') = \sum_{i=0}^n p_i \exp \left[- \sum_{j=1}^J g_j(i) y'_j \right]. \tag{14b}
\]

All the other points on the same \( x' \)-leave are no longer relevant: they are deemed statistically impossible under the prior \( p \) and observed \( x' \). The foliation therefore represents a partition of the \( \mathcal{M}_n \) into macro- and micro-worlds: Transversing between different \( x' \)-leaves are macroscopic thermodynamic processes that follow the \( y'(x') \). According to the logic of Bayesian statistics, one should use the most suitable probability frequency distribution \( \hat{v}^*(x') \) to update the prior \( p \) for the particular system with observed \( x' \). The microscopic world is still random, due to missing information, but its prior is now updated. This is Gibbs’ statistical ensemble.

With a given set of \( (g_1, \cdots, g_j) \), the \( \mathcal{M}_n \) is collapsed into \( j \)-manifold in \( \mathbb{R}^{n+1} \), which is parametrized by the \( x' \), or equivalently \( y' \). There is no uncertainty in this “macroscopic” description. For a different set of \( g \)'s and \( f' \), there will be a different \( f' \)-manifold. It will be desirable to treat different \( g \)'s through transformations. We note that even though \( \mathcal{M}_n \) is a “plane” in \( \mathbb{R}^{n+1} \), it is not a linear Euclidean space since for any \( c \neq 1, c \hat{v} \notin \mathcal{M}_n \), and neither are the \( x' \)-leaves. They are affine manifolds [26]. The locating of \( \hat{v}^*(x') \) is a highly nonlinear procedure in the space of energies.

The LFT, in terms of \( \Psi(y), \Psi(y) \), etc., enters as a powerful algebraic linear representation of the MEP procedure. The “collapse” of a holographic \( y \) to \( y' \) with missing information means simply neglecting all the extra dimensions: \( y_{j+1} = \cdots = y_n = 0 \). This is because due to the convexity of \( \Phi(\hat{v}) \), there is a one-to-one relation between \( \hat{v} \) and \( u = -\nabla \Phi \) under a proper gauge fixing. And since the constrains to MEP in (5) are all linear due to the nature of observables being random variables, each \( g \) determines a 1-dimensional linear subspace in the space of \( u \).

5. Generalized Clausius Inequality

A combination of Equations (9) and (12a) yields a Clausius’ inequality-like relation:

\[
\varphi(x') + x' \cdot y' - \psi(y') \geq 0. \tag{15}
\]

The thermodynamics equilibrium is between the observed mean value \( x' \) and its conjugate “force” \( y' \). When the equality holds, there is a relation between \( x' \) and \( y' \) which should be identified as a “the equation of state”, with \( \nabla_x \varphi = -y' \) and \( \nabla_y \psi = x' \). When the \( x' \neq \nabla_y \psi(y) \), the difference \( x' \cdot y' - \psi(y') \) can be interpreted as the nonequilibrium heat and \( \varphi \) again as the entropy; then, the inequality in (15) becomes the Clausius’ inequality.
6. Generalized Gibbs–Duhem Equation

The celebrated Gibbs–Duhem equation in classical thermodynamics is a consequence of the entropy being a Eulerian degree 1 homogeneous function. Thus, for the \( \Phi (\nu) \) in (1), we have

\[
\Phi (\nu) = \sum_{i=0}^{n} v_i \left( \frac{\partial \Phi}{\partial v_i} \right), \quad \sum_{i=0}^{n} v_i \left( \frac{\partial^2 \Phi}{\partial v_i \partial v_j} \right) = 0,
\]

\[
\sum_{j=0}^{n} \frac{\partial^2 \Phi}{\partial v_j} = \sum_{i=0}^{n} v_i \sum_{j=0}^{n} \frac{\partial u_i}{\partial v_j} \, dv_j = 0,
\]

that is, \( \sum_{i=0}^{n} v_i \, du_i = 0, \) (16)

in which we have used (2b). We identify (16) as a generalized Gibbs–Duhem equation.

7. Conclusions

The mathematical theory of probability deals with a set of elementary events \( S \), on which the probability \( p \) and random variables \( g \)'s are introduced. Applying this mathematics to the real world, each ergodic dynamical system with state space \( S \) has its own unique steady-state probability distribution which can be obtained as the \( \hat{\nu} \) from i.i.d. sampling ad infinitum.

Our present theory is to statistical inference obtaining particular \( \hat{\nu} \)'s what dynamics is to kinematics in classical mechanics [27]. The entropy function in (1) arises in this context as a measure of the quantitative relationship between the assumed, “hypothesis” \( p \) and the observed “data” \( \nu \) and \( \hat{\nu} \), as “missing information” or “surprise” [21,22]. Motivated by the analogy to the Fourier analysis, our generalized Gibbs’ theory suggests that the notion of thermo-energetics is a powerful mathematical transformation of the statistical description; \( \hat{\nu} \) and \( \nu \) are simply two representations of a same physical reality, the former being statistical while the latter thermo-energetic. \( \nu \) is to \( \hat{\nu} \) what chemical potential \( \mu \) is to particle number \( N \) in J. W. Gibbs’ chemical thermodynamics. With a fixed \( p \), the theory of probability [8] revealed a powerful, dual energetic representation for various different systems, with the same state space \( S \), in terms of their respective internal energy functions \( \nu \) [25]. This fundamental duality between counting frequency and internal energy of course has been recognized by L. Boltzmann already in 1880s, when he was developing the statistical mechanics as a foundation of classical thermodynamics under the principle of equal probability a priori. The present work shows that while the probability and statistics are fundamental as the foundation of thermodynamics, mechanics is not necessary. A similar conclusion was reached in the 1925 thesis of L. Szilard [28,29].

For sufficiently large \( N \), the probability of observing a particular \( \hat{\nu} \) is asymptotically zero except \( \hat{\nu} = p \). The significance of \( \Phi (\hat{\nu}; p) \) is to provide a “high-resolution magnifying glass” for the asymptotically small

\[
\exp \left\{ -N \Phi (\hat{\nu}; p) \right\}. \quad (17)
\]

This is known as the large deviations rate function in the modern theory of probability [8]. The entropy \( -\Phi (\hat{\nu}, p) \) is a function of both \( \hat{\nu} \) and \( p \), \( \Phi (\hat{\nu}; p) \geq 0 \) and \( \Phi (p; p) = 0 \). For a given \( p \), it views each possible \( \hat{\nu} \) from a real system as a part of an entire class of systems under a common \( p \), a meta-statistics. If one chooses the true steady-state probability \( \pi \) of a particular system to replace \( p \), then Equation (17) gives the probability distribution of the uncertainties in the measurement \( \hat{\nu} \) from \( N \) samples. The second-order Taylor expansion near \( \pi \),

\[
e^{-N\Phi (\hat{\nu}; \pi)} \simeq \exp \left[ -N \sum_{i,j=0}^{n} \frac{(\delta_i - \pi_i)(\delta_j - \pi_j)(\hat{\nu}_i - \pi_i)}{\pi_i} \right],
\]
is the central limit theorem for the statistics of counting frequency \( \hat{\nu} \), with \( \text{Var}[\hat{\nu}] = \pi_i(1 - \pi_i)/N \) and \( \text{Cov}[\hat{\nu}_i, \hat{\nu}_j] = -\pi_i \pi_j / N \). This is not the fluctuations within the \( \pi \) of the system itself. Gibbs’ theory of ensemble is about statistical measurements of a whole system; not about the individuals within.

We choose to present our theory with finite state space \( S \) for mathematical simplicity. Formal generalization to continuous state space is straightforward if mathematical rigor is not required. Beyond the finite state space, it is known that modern probability and the theory of measures encounter challenges, c.f., de Finetti’s treatment of infinite sets and the axiom of choice of nonempty subsets [20]. In addition to continuous \( \mathbb{R}^n \) [30], there are even larger Hilbert spaces of functions on \( \mathbb{R}^n \) and/or von Neumann algebra of operators acting on a Hilbert space.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: I thank Jin Feng, Weishi Liu, Zhang-Ju Liu, Bing Miao, Zhongmin Shen, Xiang Tang, Yong-Shi Wu, and particularly Jun Zhang, for many helpful discussions, and the support from Olga Jung Wan Endowed Professorship.

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

Abbreviations

The following abbreviations are used in this manuscript:

LFT  Legendre–Fenchel transform
FT  Fourier transform

Appendix A. Statistical Counting ad Infinitum

In this section, we provide the mathematical reasoning for stating “entropy provides the infinitesimal probability for an observed frequency \( \hat{\nu} \) with respect to a probability prior \( p \), “it characterizes the relationship between \( \nu \) and \( p \) in a sampling process”, and the origin of Legendre–Fenchel transform in entropy analysis. The counting of independent and identically distributed samples with state space \( S = \{0, \cdots, n\} \) yields \( \nu = (\nu_0, \cdots, \nu_n) \), a \((n + 1)\)-tuple of non-negative integers. We call all the \( \nu \) with \( N = \nu_0 + \cdots + \nu_n \) a simplex for counting. The simplex for counting grows with \( N \), which we shall identify as “time”. With a given prior probability \( p = (p_0, \cdots, p_n) \) on \( S \), statistical counting is a Markov process on a growing simplex, with probability:

\[
P^{(N+1)}(\nu) = \sum_{k=0}^{n} p_k P^{(N)}(\nu - \delta_k), \tag{A1}
\]

in which \( \delta_k = (0, \cdots, 1, \cdots, 0) \) is the unit vector for the \( k^{th} \) component. One can easily verify that

\[
P^{(N)}(\nu) = \frac{N!}{\nu_0! \cdots \nu_n!} p_0^{\nu_0} \cdots p_n^{\nu_n}
\]

is a solution to (A1).

One is interested in the limit of counting \( \text{ad infinitum} \), when all the \( \nu_i \)s are expected to tend to infinity as \( N \to \infty \). On the increasing simplex for \( \nu \), the probability \( P^{(N)}(\nu) \to 0 \). However, the properly normalized \( \hat{\nu} = \nu/N \) converges, and \( P^{(N)} \) as a function of the \( \hat{\nu} \) becomes sharper and sharper, concentrated around \( \hat{\nu}^* = p \). To more precisely characterize this limiting situation, one introduces counting frequency \( \hat{\nu} = \nu/N \). The space of \( \hat{\nu} \)s then is called a probability simplex \( \mathcal{M}_n \); Equation (A1) then becomes
\[ p^{(N+1)}(\hat{\nu}) = \sum_{k=0}^{n} p_k p^{(N)} \left( \frac{n+1}{N} \hat{\nu}_i - \frac{1}{N} \delta_{ik} \right). \]  

(A2)

Its limit is a Dirac-\(\delta\) function: \(p^{(\infty)} = 0\) for all \(\hat{\nu} \neq \nu\), and \(p^{(\infty)} = \infty \) at \(\hat{\nu} = \nu\). However, “a higher order” infinitesimal analysis shows that [8]

\[ \lim_{N \to \infty} \frac{1}{N} \ln p^{(N)}(\hat{\nu}) = -\sum_{i=0}^{n} \hat{\nu}_i \ln \left( \frac{\hat{\nu}_i}{p_i} \right) = -\Phi(\hat{\nu}). \]  

(A3)

It is clear that entropy function \(-\Phi(\nu)\) represents the infinitesimal prior probability \(e^{-N\Phi(\nu)}\) on \(\mathcal{M}_u\). For two \(\nu\)’s with different entropy values, \(\Phi(\nu)\) and \(\Phi(\nu')\), their probabilities \(p^{(\nu)}(\nu)/p^{(\nu')}(\nu') = 0\) if \(\Phi(\nu) > \Phi(\nu')\). This is the origin of the maximum entropy principle (MEP).

To understand the limit \(p^{(N)}(\nu) \to 0\), one can also introduce the probability generating function [8]:

\[ W^{(N)}(u) = \sum_{\nu} p^{(N)}(\nu) e^{-u \nu}, \]  

(A4)

in which \(u \cdot v = u_0 v_0 + \cdots + u_n v_n\). Then, Equation (A1) becomes

\[
\begin{align*}
W^{(N+1)}(u) &= \sum_{\nu} p^{(N+1)}(\nu) e^{-u \nu} \\
&= \sum_{k=0}^{n} p_k e^{-u \delta_k} \sum_{\nu} p^{(N)}(\nu - \delta_k) e^{-u (\nu - \delta_k)} \\
&= W^{(N)}(u) e^{-\Psi(u)},
\end{align*}
\]

where \(\Psi(u) = -\ln \sum_{k=0}^{n} p_k e^{-u_k}\).  

(A5)

The free energy function \(\Psi\) is meaningful for all finite \(N\). This is why the partition function is valid even for small systems in Gibbs’ theory of ensembles [13]. The Legendre–Fenchel transform of \(\Psi(u)\) is precisely the the right-hand side of (A3):

\[
\begin{align*}
\inf_u \left\{ u \cdot v - \Psi(u) \right\} &= \inf_u \left\{ \sum_{i=0}^{n} v_i \ln e^{u_i} + \ln \sum_{k=0}^{n} p_k e^{-u_k} \right\} \\
&= \inf_u \left\{ -\sum_{i=0}^{n} v_i \ln \left[ \frac{e^{-u_i}}{\sum_{k=0}^{n} p_k e^{-u_k}} \right] \right\} = -\sum_{i=0}^{n} v_i \ln \left( \frac{\hat{\nu}_i}{p_i} \right),
\end{align*}
\]

(A6)

in which the optimal \(e^{-u_i} \propto v_i/p_i\). Legendre–Fenchel transform arises in the limit of \(N \to \infty\) through the Laplace’s method of evaluating asymptotic integrals, or the related Darwin–Fowler method of maximum term.

The analysis in this Appendix suggests that a proper interpretation of \(p = (p_0, \cdots, p_n)\) in Equation (A1) is not as an intrinsic property, for example, the generative model of data statistics, rather it should be interpreted as a choice of a “gauge” in terms of which a set of counting data is represented: each particular set of data \(ad \ infinitum\) is represented by the energy function \(u\), not \(p\), and the \(\hat{\nu}\) is gauge invariant via the Boltzmann relation \(\hat{\nu}_i \propto p_i e^{-u_i}\)—this yields an i.i.d. generative model. Probability is not for generative models, it is for analyzing empirical measurements on random variables.

References
1. Førland, K.S.; Førland, T.; Kjelstrup, S. Irreversible Thermodynamics: Theory and Applications; John Wiley & Sons: Chichester, UK, 1988.
2. Kjelstrup, S.; Bedaux, D. Non-Equilibrium Thermodynamics of Heterogeneous Systems; Series on Advances in Statistical Mechanics; World Scientific: Singapore, 2008; Volume 16. [CrossRef]
3. Bedeaux, D.; Kjelstrup, S.; Schnell, S.K. *Nanothermodynamics Theory and Applications*; World Scientific: Singapore, 2023. [CrossRef]
4. Guggenheim, E.A. *Modern Thermodynamics by the Methods of Willard Gibbs*; Methuen & Co.: New York, NY, USA, 1933.
5. Hill, T.L. A different approach to nanothermodynamics. *Nano Lett.* 2001, 1, 273–275. [CrossRef]
6. Khinchin, A.Y. *Mathematical Foundations of Statistical Mechanics*; Dover: New York, NY, USA, 1949.
7. Touchette, H. The large deviation approach to statistical mechanics. *Phys. Rep.* 2009, 478, 1–69. [CrossRef]
8. Dembo, A.; Zeitouni, O. *Large Deviations Techniques and Applications*, 2nd ed.; Springer: New York, NY, USA, 1998. [CrossRef]
9. Anderson, P.W. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science* 1972, 177, 393–396. [CrossRef] [PubMed]
10. Alm, J.F.; Walker, J.S. Time-frequency analysis of musical instruments. *SIAM Rev.* 2002, 44, 457–476. [CrossRef]
11. Fourier, J.B.J. *The Analytic Theory of Heat*; Freeman, A., Translator; Cambridge University Press: London, UK, 1878. [CrossRef]
12. Truesdell, C. *Rational Thermodynamics*; Springer: New York, NY, USA, 1984. [CrossRef]
13. Lu, Z.; Qian, H. Emergence and breaking of duality symmetry in thermodynamic behavior: Repeated measurements and macroscopic limit. *Phys. Rev. Lett.* 2022, 128, 150603. [CrossRef] [PubMed]
14. Galteland, O.; Bering, E.; Kristiansen, K.; Bedeaux, D.; Kjelstrup, S. Legendre-Fenchel transforms capture layering transitions in porous media. *Nanoscale Adv.* 2022, 4, 2660–2670. [CrossRef] [PubMed]
15. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.
16. Qian, M.; Xie, J.S.; Zhu, S. *Smooth Ergodic Theory for Endomorphisms*; Lecture Notes in Mathematics; Springer: Berlin, Germany, 2009; Volume 178. [CrossRef]
17. Dorfman, J.R. *An Introduction to Chaos in Nonequilibrium Statistical Mechanics*; Cambridge Lect. Notes in Phys.; Cambridge University Press: London, UK, 1999. [CrossRef]
18. Mackey, M.C. The dynamic origin of increasing entropy. *Rev. Mod. Phys.* 1989, 61, 981–1015. [CrossRef]
19. Qian, H.; Ge, H. *Stochastic Chemical Reaction Systems in Biology*; Lect. Notes on Math. Modelling in the Life Sci.; Springer Nature: Cham, Switzerland, 2021. [CrossRef]
20. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: London, UK, 2003.
21. Levine, R.D. Information theory approach to molecular reaction dynamics. *Annu. Rev. Phys. Chem.* 1978, 29, 59–92. [CrossRef]
22. Ben-Naim, A. *A Farewell to Entropy: Statistical Thermodynamics Based on Information*; World Scientific: Singapore, 2008. [CrossRef]
23. Callen, H.B. *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed.; Wiley: New York, NY, USA, 1991.
24. Commons, J.; Yang, Y.J.; Qian, H. Duality symmetry, two entropy functions, and an eigenvalue problem in Gibbs’ theory. *arXiv* 2021. [CrossRef]
25. Qian, H. Statistical chemical thermodynamics and energetic behavior of counting: Gibbs’ theory revisited. *J. Chem. Theory Comput.* 2022, 18, 6421–6436. [CrossRef]
26. Hong, L.; Qian, H.; Thompson, L.F. Representations and divergences in the space of probability measures and stochastic thermodynamics. *J. Comput. Appl. Math.* 2020, 376, 112842. [CrossRef]
27. Goldstein, H. *Classical Mechanics*; Addison-Wesley: New York, NY, USA, 1951.
28. Szilard, L. Über die ausdehnung der phänomenologischen thermodynamik auf die schwankungsserscheinungen. *Z. Physik.* 1925, 32, 753–788. [CrossRef]
29. Mandelbrot, B. On the derivation of statistical thermodynamics from purely phenomenological principles. *J. Math. Phys.* 1964, 5, 164–171. [CrossRef]
30. Miao, B.; Qian, H.; Wu, Y.S. Emergence of Newtonian deterministic causality from stochastic motions in continuous space and time. *arXiv* 2024. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.