

Duality symmetry, two entropy functions, and an eigenvalue problem in generalized Gibbs' theory

Jeffrey Commons, Ying-Jen Yang(杨颖任), and Hong Qian(钱纮)
Citation: Chin. Phys. B, 2025, 34 (**8**): 080201. DOI: 10.1088/1674-1056/adbed7
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Programming guide for solving constraint satisfaction problems with tensor networks

Xuanzhao Gao(高煊钊), Xiaofeng Li(李晓锋), and Jinguo Liu(刘金国)
Chin. Phys. B, 2025, 34 (**5**): 050201. DOI: 10.1088/1674-1056/adbee2

Self-similarity of multilayer networks

Bing Wang(王冰), Huizhi Yu(于蕙芷), and Daijun Wei(魏代俊)
Chin. Phys. B, 2025, 34 (**1**): 010202. DOI: 10.1088/1674-1056/ad8a47

On fractional discrete financial system: Bifurcation, chaos, and control

Louiza Diabi, Adel Ouannas, Amel Hioual, Shaher Momani, and Abderrahmane Abbes
Chin. Phys. B, 2024, 33 (**10**): 100201. DOI: 10.1088/1674-1056/ad5d96

Quintessence anisotropic stellar models in quadratic and Born-Infeld modified teleparallel Rastall gravity

Allah Ditta, Tiecheng Xia(夏铁成), Irfan Mahmood, and Asif Mahmood
Chin. Phys. B, 2024, 33 (**3**): 030204. DOI: 10.1088/1674-1056/acfaf9

An incommensurate fractional discrete macroeconomic system: Bifurcation, chaos, and complexity

Abderrahmane Abbes, Adel Ouannas, and Nabil Shawagfeh
Chin. Phys. B, 2023, 32 (**3**): 030203. DOI: 10.1088/1674-1056/ac7296

SPECIAL TOPIC — A celebration of the 90th Anniversary of the Birth of Bolin Hao

Duality symmetry, two entropy functions, and an eigenvalue problem in generalized Gibbs' theory

Jeffrey Commons[1], Ying-Jen Yang(杨颖任)[2,3], and Hong Qian(钱纮)[2,†]

[1] *Department of Physics, University of Washington, Seattle, WA* 98195-1560, *USA*
[2] *Department of Applied Mathematics, University of Washington, Seattle, WA* 98195-3925, *USA*
[3] *Laufer Center for Physical and Quantitative Biology, Stony Brook University, New York* 11794, *USA*

(Received 30 June 2024; revised manuscript received 2 March 2025; accepted manuscript online 11 March 2025)

We generalize the convex duality symmetry in Gibbs' statistical ensemble formulation, between the Gibbs entropy $\varphi_{V,N}(E)$ as a function of mean internal energy E and Massieu's free entropy $\Psi_{V,N}(\beta)$ as a function of inverse temperature β. The duality in terms of Legendre–Fenchel transform tells us that Gibbs' thermodynamic entropy is to the law of large numbers (LLN) for arithmetic sample mean values what Shannon's information entropy is to the LLN for empirical counting frequencies in independent and identically distributed data. Proceeding with the same mathematical logic, we identify the energy of the state $\{u_i\}$ as the conjugate variable to the counting of statistical occurrence $\{m_i\}$ and find a Hamilton–Jacobi equation for the Shannon entropy analogous to an equation of state in thermodynamics. An eigenvalue problem that is reminiscent of certain features in quantum mechanics arises in the entropy theory of statistical counting frequencies of Markov correlated data.

Keywords: counting statistics, equation of state, entropy, large deviations, law of large numbers

PACS: 02.50.Cw, 05.70.–a, 82.60.–s, 89.70.–a

DOI: 10.1088/1674-1056/adbed7 **CSTR:** 32038.14.CPB.adbed7

1. Introduction

In recent years, through the development in stochastic thermodynamics and the applications of the mathematical theory of large deviations to the asymptotic behavior of various forms of the law of large numbers (LLN), it has been recognized that "the essence of a thermodynamic description is not found in its connection to conservation laws, microscopic reversibility, or the equilibrium state relation they entail, despite the central role those play"; see Refs. [1–3] and the references cited within. In addition to derivations of fluctuation relations in connection to nonequilibrium entropy productions, more recent results on the unification of stochastic chemical kinetics and Gibbs' thermochemistry,[4,5] as well as the probabilistic elucidation of T. L. Hill's thermodynamics of small systems (1963),[6,7] have all further substantiated the above claim.

The large deviation theory (LDT) from the mathematics of probability[8,9] provides the analytic underpinning for understanding a range of law-like behaviors of statistical observables, of which the arithmetic mean values and normalized empirical counting frequencies (proportions) are the two most important types. LDT is an emergent behavior that arises from the mathematical limit of infinitely many data whose statistics are without randomness;[10] in fact we shall show that the Gibbsian statistical thermodynamics can be framed, independent from mechanics or any mechanistic underlying, as the asymptotic statistical laws for counting frequencies together with associated maximum entropy principle via Lagrangian duality.[3,11] The mathematical analysis reveals an important duality structure between two functions, a large deviation rate function (LDRF) and a cumulant generating function (CGF). In statistical mechanics, they correspond precisely to Gibbs' (neg) entropy as a function of all extensive variables, $S(E,V,N)$, and the Massieu entropy as a function of all intensive, conjugate variables, $\Psi(T,p,\mu)$.[7] The latter are often called "free energy" in mathematical literature, where neg-entropy and entropy are referred to indiscriminately. When $S(E,V,N)$ is an Eulerian homogeneous function of degree 1, the existence of its convex dual $\Psi(T,p,\mu)$ requires the more refined modern formulation of Legendre–Fenchel transform[12] in terms of indicator function and support function duality.[13]

LDT clearly resolves the contradistinction between the two best-known entropies, that of Gibbs and that of Shannon.[14] Although writings on the subject are already vast, the present work gives a more coherent presentation showing that the Gibbs entropy, defined via temperature derivative of canonical partition function as in standard textbooks, is to the LLN for arithmetic mean values what the Shannon entropy is to the LLN for empirical counting frequencies. They are the subjects of Crámer's theorem and the Sanov theorem in mathematics, respectively.[9]

With this newfound perspective and by explicating the

†Corresponding author. E-mail: hqian@uw.edu

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

duality for the Shannon–Sanov relative entropy,[15,16] also widely called Kullback–Leibler divergence, we show that the conjugate variables to "countings of statistical occurrences" $\{m_i\}$ are a set of "energy parameters" in $k_{\rm B}T$ units, $\{u_i\}$. The duality turns out to be precisely Gibbs' statistical thermodynamics. His theory for chemical thermodynamics with a few macroscopic variables and his formulation for statistical mechanics in terms of the entire phase space $\mathbb{R}^N$ with N degrees of freedom, thus, are unified. Via Boltzmann's relation (see Eq. (20) below) there is an equivalence principle between the statistical kinematics of counting and the energetic description of states. We further consider repeated measurements on a time-correlated Markov process and find a novel eigenvalue problem for the square root of empirical frequency $\{\sqrt{v_i}\}$.

In recent years,[7] it has been demonstrated that the application of the mathematics of Legendre–Fenchel duality[12] is at the core of Gibbs' modern thermodynamics for macroscopic variables.[17,18] Following this logic,[3] it was reported that even Gibbs' theory of statistical mechanics, in terms of an energy function and its logarithmic partition function as free energy, is mathematically derivable by the methods of Gibbs. The present work is structured as follows. The previous results are summarized in Section 2. In Section 3, in parallel to counting states that correspond to statistical thermodynamics, we consider counting for consecutive states, e.g., statistics on discrete-time kinetics, and introduce a new entropy function. We further derive an entropy function for continuous-time counting. This novel entropy function in Eq. (32) has a quadratic form of the square roots of probabilities, rather than based on logarithmic probability as the Shannon entropy and its variations. Then in Eq. (33) we further show an eigenvalue problem that is associated with the Legendre–Fenchel transform (LFT) of the entropy function. The possible physical interpretations of the results and additional discussion are given in Section 4.

2. Duality symmetry in Gibbs' theory and its generalization

2.1. Duality symmetry between $\psi_{V,N}(\beta)$ and $\varphi_{V,N}(E)$

It is important to recall that as a thermodynamic potential function, the Gibbs entropy is a function of an extensive variable E, the internal energy, and the Helmholtz free energy or Massieu free entropy is a function of the intensive conjugate variable T.[18] There is a convex duality between these two functions[12] as follows.

Consider a mechanical system consisting of N degrees of freedom, $\boldsymbol{x} = (x_1, \ldots, x_N)$, in a container with volume V under temperature $T = (k_{\rm B}\beta)^{-1}$, with a potential energy function $U_N(\boldsymbol{x})$. The logarithm of the canonical partition function

$$\Psi_{V,N}(\beta) = k_{\rm B} \log \int_{V^N} {\rm e}^{-\beta U_N(\boldsymbol{x})} \, {\rm d}\boldsymbol{x} \tag{1}$$

has been known as Massieu's free entropy. See Ref. [7] for a very recent discussion. Then the equilibrium entropy of the statistical mechanical system according to Gibbs' theory is

$$S(\beta) = \frac{\partial[\beta^{-1}\Psi_{V,N}(\beta)]}{\partial \beta^{-1}} = \Psi_{V,N}(\beta) - \beta \frac{\partial \Psi_{V,N}}{\partial \beta}, \tag{2}$$

which can be expressed in terms of LFT from convex analysis[12]

$$\varphi_{V,N}(E) \equiv -S\big(\beta(E)\big) = -\inf_{\beta}\Big\{\beta E + \Psi_{V,N}(\beta)\Big\} \tag{3a}$$

$$= \begin{cases} -\beta E - \Psi_{V,N}(\beta), \\ E(\beta) = -\partial \Psi_{V,N}(\beta)/\partial \beta, \end{cases} \tag{3b}$$

in which relation $E(\beta)$ is monotonic since $\Psi_{V,N}$ is a convex function of β, therefore $\beta(E)$ is well defined. Equation (3b) is a parametric expression for the function $\varphi_{V,N}(E)$ in terms of the parameter β. And since $\Psi_{V,N}(\beta)$ is a convex function of β, Eq. (3) has a dual expression

$$\Psi_{V,N}(\beta) = -\inf_{E}\Big\{\beta E + \varphi_{V,N}(E)\Big\}, \tag{4}$$

and a corresponding pair of differential equations relating φ and Ψ as thermodynamic potential functions for their respective independent variables, with corresponding conjugate variables β and E:

$${\rm d}\varphi_{V,N}(E) = \left(\frac{{\rm d}\varphi_{V,N}(E)}{{\rm d}E}\right){\rm d}E = -\beta(E)\,{\rm d}E, \tag{5a}$$

$${\rm d}\Psi_{V,N}(\beta) = \left(\frac{{\rm d}\Psi_{V,N}(\beta)}{{\rm d}\beta}\right){\rm d}\beta = -E(\beta)\,{\rm d}\beta. \tag{5b}$$

The duality in Eq. (5) can be further extended to intensive conjugate variables p and μ which correspond to extensive variables V and N. This leads to a duality symmetry between the fundamental thermodynamic relation and the Hill–Gibbs–Duhem (HGD) equation in nanothermodynamics,[7] respectively:

$${\rm d}S = \beta\,{\rm d}E + (\beta p)\,{\rm d}V - (\beta\mu)\,{\rm d}N, \tag{6a}$$

$${\rm d}\big(\beta\mathscr{E}\big) = E\,{\rm d}\beta + V\,{\rm d}\big(\beta p\big) - N{\rm d}\big(\beta\mu\big), \tag{6b}$$

where S, E, V and N are entropy, internal energy, volume and particle number, and inverse temperature β, pressure p, chemical potential μ are their conjugate variables, and $\mathscr{E}(\beta, p, \mu)$ is the novel subdivision potential introduced in nanothermodynamics, as the LFT of non-extensive $S(E, V, N)$.

In the classical macroscopic thermodynamic limit, the HGD equation is reduced to the Gibbs–Duhem equation, with the duality symmetry broken: When $S = \varphi_{V,N}(E)$ is an Eulerian homogeneous degree 1 function of E, V, N, its formal LFT $\beta\mathscr{E} = \psi(\beta, p, \mu) = \infty$ or 0 divided by a hyper-surface that yields the equation of state (EoS):

$$\psi(\beta, p, \mu) = -\inf_{E,V,N}\Big\{\beta\big(E + pV - \mu N\big) + \varphi_{V,N}(E)\Big\}$$

$$= \mu N - \sup_V \left\{ pV - \beta^{-1}\Psi_{V,N}(\beta) \right\} = 0. \quad (7)$$

On the hyper-surface, the μ is a function of p and β; the $\psi(\beta, p, \mu)$ vanishes. In the familiar standard thermodynamics notations, $\Psi = -F/T = -(E-TS)/T$, where $1/T = \partial S/\partial E$ and $\mu = G/N = (pV+F)/N$ with $p = -\partial F/\partial V$. They correspond to Eqs. (4) and (7), with the infimum and supremum carried out by differentiating convex functions.

2.2. The Gibbs entropy and the Shannon–Sanov entropy in current understanding

For a given probability distribution over a set of n discrete events with probability mass function p_i, $1 \le i \le n$, Shannon's entropy for empirical frequencies $\{\nu_i\}$ is a special case of the more general relative entropy[15,16]

$$H(\{\nu_i\}) = \sum_{i=1}^{n} \nu_i \log \frac{\nu_i}{p_i} \quad \text{or } H[f] = \int_{\mathbb{R}} f(x) \log \frac{f(x)}{p(x)} \mathrm{d}x. \quad (8)$$

Shannon's original formula is taken relative to a uniform probability: $-H(\nu) = -\sum_{i=1}^{n} \nu_i \log \nu_i - \log n$. There is an extra, non-consequential constant $-\log n$. The first term is interpreted as the "missing information" w.r.t. the maximum information $-\log n$.

According to standard textbooks on equilibrium statistical mechanics,[19] the Gibbs entropy is defined by starting from a given density of states in terms of energy $\Omega(E)$, $u \in \mathbb{R}$,

$$F(\beta) = -\beta^{-1} \log \int_{\mathbb{R}} \Omega(E) \mathrm{e}^{-\beta E} \mathrm{d}E, \; S = -\frac{\partial F(\beta)}{\partial (1/\beta)}, \quad (9)$$

where S is the Gibbs entropy. F is the Helmholtz free energy and $\beta = (k_{\mathrm{B}}T)^{-1}$ is the inverse temperature. A canonical distribution is

$$f(E) = \Omega(E) \mathrm{e}^{-\beta E} \left(\int_{\mathbb{R}} \Omega(E) \mathrm{e}^{-\beta E} \mathrm{d}E \right)^{-1},$$

from which one has the well-known equation

$$S = -\frac{\partial F(\beta)}{\partial (1/\beta)} = -\int_{\mathbb{R}} f(E) \log \left(\frac{f(E)}{\Omega(E)} \right) \mathrm{d}E. \quad (10)$$

If one identifies $\Omega(E)$ with the uniform density $1/N$ and integration w.r.t. E with summation over n; then the Gibbs entropy and the Shannon entropy are considered to be the same; this is self-evident from the two expressions (8) and (10).

2.3. Beyond the current understanding

In Eqs. (3) and (4), with fixed V and N, $\varphi_{V,N}(E)$ as a function of E is the LFT of $\Psi_{V,N}(\beta)$ as a function of β, whose LFT in return is $\varphi_{V,N}(E)$. Even though both entropy and free energy are state functions in thermodynamics and their independent variables can be arbitrarily chosen as E or β, according to the mathematical theory of LFT, specifying the independent variables of an entropy function matters. This is also clear from the writing in Eqs. (2) and (3a). In Eq. (8) it is a function of $\nu_i = m_i/M$, or functional of $f(x)$; the independent variables here are empirical frequency distributions. Entropy in thermodynamics is a function of macroscopic variables. In the simplest case, it is a function of E, V, and N. The latter is obtained from the former via maximum entropy principle under constraints.[2,3,11] While one has found an agreement between the mathematical expressions for the two entropy functions, why should the mathematical function in Eq. (8) be minimized? Though originating from thermodynamics, the principle of maximum entropy must have a deeper root in statistics; the answer has to be found in the mathematical logic of probability.

There is a new principle that calls for revisiting the two entropy functions:[3] *The entropy function arises as the asymptotic exponent for a law of large numbers*. There are many different laws of large numbers, thus there are many different entropy functions, c.f. Eq. (26) below. Gibbs' and Shannon's are merely two special ones for the two most widely studied LLNs. We shall consider two i.i.d. sampling problems: (i) a discrete random event with probability mass function p_k, $1 \le k \le n$, and (ii) a continuous real-valued random variable X with probability density function $\Omega(x)$.

Shannon-Sanov's relative entropy as a LDRF — For the discrete problem, let us assume that within total M i.i.d. samples, there are m_k observations of the k^{th} event: $m_1 + \cdots + m_n = M$. It is clear that if one carries out another M observations, one would obtain another set of $\{m_k; 1 \le k \le n\}$. So the m_k's are themselves random variables with a joint probability distribution

$$\frac{M!}{m_1! \ldots m_n!} p_1^{m_1} \cdots p_n^{m_n}. \quad (11)$$

As $M \to \infty$, the normalized counting frequency $\nu_k := \frac{m_k}{M} \to p_k$: The proportion of observing the k^{th} event approaches its probability. This is known as Borel's LLN.

The large deviation principle says that the distribution in expression (11), as $M \to \infty$, has an asymptotic expression $\mathrm{e}^{-M\varphi(\nu_1,\ldots,\nu_n)}$, where $\varphi(\nu_1, \ldots, \nu_n)$ is exactly the $H(\{\nu_i\})$ in Eq. (8).

The Gibbs entropy is also a LDRF — Now for the continuous problem, let $X_1, X_2, \ldots, X_M$ be M i.i.d. samples of a random variable X, with probability density function $\Omega(x)$. Then the LLN for the mean value says that

$$\overline{X}_M := \frac{X_1 + \ldots + X_M}{M} \to \mathbb{E}[X] = \int_{\mathbb{R}} x\Omega(x) \mathrm{d}x,$$

as $M \to \infty$. It is clear that if one obtains another M samples, the $\overline{X}_M$ will be different. So $\overline{X}_M$ itself is a random variable with a probability density function $f_{\overline{X}_M}(x)$.

Again, the LDT says that as $M \to \infty$, $f_{\overline{X}_M}(x)$ has an asymptotic expression $\mathrm{e}^{M\phi(x)}$. To compute the exponent $\phi(x)$,

we employ the method of CGF,

$$\Lambda(\beta) = \log \int_{\mathbb{R}} \Omega(x)\, \mathrm{e}^{-\beta x} \mathrm{d}x, \tag{12}$$

and note that the CGF of $\overline{X}_M$ is $\Lambda_{\overline{X}_M}(\beta) = M\Lambda(\beta/M)$. Then

$$\Lambda(\beta) = \frac{1}{M}\Lambda_{\overline{X}_M}(M\beta) = \sup_x \{\phi(x) - \beta x\}, \tag{13}$$

which implies

$$\phi(x) = \inf_\beta \{\beta x + \Lambda(\beta)\}. \tag{14}$$

Now to carry out the optimization problem in Eq. (14), let us use calculus to find the optimal β^*: $x + \Lambda'(\beta) = 0$. Therefore,

$$\begin{cases} \phi &= -\beta\Lambda'(\beta) + \Lambda(\beta) = \dfrac{\partial}{\partial(1/\beta)}\left[\beta^{-1}\Lambda(\beta)\right], \\ x &= -\dfrac{\partial}{\partial\beta}\Lambda(\beta). \end{cases} \tag{15}$$

Equation (15) gives the function $\phi(x)$ in a parametric form. We see that if we identify $-\beta^{-1}\Lambda(\beta)$ as $F(\beta)$, then the ϕ in Eq. (15) is the Gibbs entropy, and x is the internal energy.

Therefore, as the asymptotic exponent for the probability of large deviations, the Shannon entropy is to Borel's LLN for empirical counting frequencies what the Gibbs entropy is to the standard LLN for arithmetic mean values.

Gibbs–Kirkwood potential energy and the dual to the Shannon–Sanov entropy — Through Legendre transforms, the theory of classical thermodynamics introduces intensive conjugate variables to extensive quantities, such as β to internal energy U, p to volume V, and μ to particle number N. The same conjugate variables enter Gibbs' theory of statistical mechanics via Laplace transforms of various partition functions corresponding to extensive variables. The result from the previous sections shows clearly that a duality structure between the LDRF and its CGF plays a central role in a new understanding of Gibbs' and Shannon's entropy functions.

By the same logic the following question immediately arises: What is the corresponding CGF for Shannon's relative entropy, i.e., Sanov's LDRF in Eq. (8)? We carry out its LFT using u_i to denote the conjugate variable of ν_i under the normalization constraint $\sum_i^n \nu_i = 1$:

$$\Xi(\{u_i\}) = \inf_{\{\nu_i\}} \left\{ \sum_{i=1}^n u_i\nu_i + H(\{\nu_i\}) \right\} \tag{16}$$

$$= \sum_{i=1}^n \nu_i^* \log\left(\frac{\nu_i^*}{p_i\, \mathrm{e}^{-u_i}}\right) = \lambda = -\log \sum_{i=1}^n p_i\, \mathrm{e}^{-u_i}, \tag{17}$$

in which λ appears as a Lagrange multiplier for the constrained optimization $\nu_i^* = p_i\, \mathrm{e}^{-u_i+\lambda}$. For fixed u_i, the ν_i are determined from $u_i = -\partial H(\{\nu_i\})/\partial\nu_i + \lambda$. The additive constant λ to the u_i's represents a gauge freedom, a feature that is absent in the traditional Legendre transform. Equation (16) has the exact form of Helmholtz free energy from standard statistical mechanics when a set of "energy" parameters in $k_\mathrm{B}T$ units $\{u_i\}$ is given:

$$\Xi(\{u_i\}) = \underbrace{\sum_{i=1}^n \nu_i^* u_i}_{\text{mean internal energy}} \underbrace{- \left(-\sum_{i=1}^n \nu_i^* \log\frac{\nu_i^*}{p_i} \right)}_{\text{entropy}}. \tag{18}$$

More precisely, the mean internal energy term in Eq. (18) should be understood as the mean value of Kirkwood's potential of mean force u_i.[20] The conjugate variables u_i therefore are interpreted as a "potential function" that represents the statistical deviations of the empirical frequencies $\{\nu_i\}$ away from the true probability $\{p_i\}$ in the large-deviation scenario.

We shall call the set of $\{u_i\}$ introduced in Eq. (16), the conjugate variables to the counting frequencies $\{\nu_i\}$, Gibbs–Kirkwood potential energies; they come with a "natural" additive constant λ due to the normalization of frequencies $\{\nu_i\}$. One then recovers the duality

$$\mathrm{d}H(\{\nu_i\}) = \sum_{i=1}^n \left(\frac{\partial H}{\partial \nu_i}\right) \mathrm{d}\nu_i = -\sum_{i=1}^n u_i(\{\nu_i\})\, \mathrm{d}\nu_i, \tag{19a}$$

$$\mathrm{d}\Xi(\{u_i\}) = \sum_{i=1}^n \left(\frac{\partial \Xi}{\partial u_i}\right) \mathrm{d}u_i = \sum_{i=1}^n \nu_i(\{u_i\})\, \mathrm{d}u_i, \tag{19b}$$

where

$$u_i(\{\nu_i\}) = -\log\left(\frac{\nu_i}{p_i}\right) + \lambda. \tag{20}$$

The Massieu free entropy as a function of β is the LFT of the Gibbs entropy as a function of energy; the Helmholtz free energy as a function of $\{u_i\}$ is the LFT of Shannon–Sanov entropy as a function of normalized frequencies $\{\nu_i\}$.

3. Entropy analysis in thermodynamics and LDT for Markov chain

To compare and contrast Gibbsian statistical thermodynamics with LDT, one recognizes that the concept of an extensive quantity in the former[18] should be identified with the empirical sum of M repeated measurements on a random variable in the latter: $X_1 + \cdots + X_M$. The empirical sum scales with the total number of samples M; furthermore entropy in thermodynamics is actually M times the LDRF ϕ, which is itself a function of empirical sum divided by M, $\overline{X}_M = (X_1 + \cdots + X_M)/M$. Therefore the large deviation principle in probability actually corresponds to the existence of a macroscopic entropy function in Gibbs' theory.

3.1. Thermodynamic entropy theory of statistical counting

Putting this realization into practice, we consider the "extensive" entropy function of non-normalized counting of states $\{m_i\}$. Then corresponding to the Shannon–Sanov entropy is $S(\{m_i\}) = -MH(\{m_i/M\})$, where $M = m_1 + \cdots + m_n$: It has

$S(\{\alpha m_i\}) = \alpha S(\{m_i\})$. The Legendre–Fenchel transform of the homogeneous $S(\{m_i\})$ yields a singular function

$$\begin{aligned}
&\inf_{\{m_i\}} \left\{ \sum_{i=1}^{n} m_i u_i - S(\{m_i\}) \right\} \\
&= \inf_{\{m_i\}} \left\{ \sum_{i=1}^{n} m_i u_i + \sum_{i=1}^{n} m_i \log \frac{m_i}{(m_1+\cdots+m_n)p_i} \right\} \\
&= \inf_{M} \left\{ M \inf_{\{\nu_i\}} \left\{ \sum_{i=1}^{n} \nu_i u_i + \sum_{i=1}^{n} \nu_i \log \frac{\nu_i}{p_i} \right\} \right\} \\
&= \inf_{M} \left\{ -M \log \sum_{i=1}^{n} p_i \mathrm{e}^{-u_i} \right\} \\
&= \begin{cases} 0 & \sum_{i=1}^{n} p_i \mathrm{e}^{-u_i} = 1, \\ -\infty & \sum_{i=1}^{n} p_i \mathrm{e}^{-u_i} \neq 1, \end{cases}
\end{aligned} \tag{21}$$

in which the n intensive conjugate variables u_i satisfy an equation of state (EoS)

$$\sum_{i=1}^{n} p_i \mathrm{e}^{-u_i} = 1. \tag{22}$$

The u's are related to the extensive m's:

$$\begin{aligned}
u_i &= -\log \frac{m_i}{(m_1+\cdots+m_n)p_i}, \\
m_i &= \frac{M p_i \mathrm{e}^{-u_i}}{\sum_{i=1}^{n} p_i \mathrm{e}^{-u_i}},
\end{aligned} \tag{23}$$

and there is a Gibbs–Duhem equation, as the differential form of the EoS in Eq. (22):

$$\begin{aligned}
\sum_{j=1}^{n} m_j(\{u_i\})\,\mathrm{d}u_j &= \frac{\sum_{j=1}^{n} p_j \mathrm{e}^{-u_j}\,\mathrm{d}u_j}{\sum_{i=1}^{n} p_i \mathrm{e}^{-u_i}} \\
&= -\mathrm{d}\left(\log \sum_{i=1}^{n} p_i \mathrm{e}^{-u_i} \right) = 0.
\end{aligned} \tag{24}$$

It is clear that the Lagrange multiplier λ in Eq. (17) is conjugate to M. We observe that by employing a complete convex analysis one actually fixed the $\lambda = 0$ or $\mathrm{e}^{\lambda} = 1$ in Eq. (22). In the theory of probability and statistical data analysis, the "large number" M is given and known. However, in the thermodynamics of a system in continuous space and time the M is unknown;[13] this leaves the to-be-determined free energy of the entire system as the logarithm of a partition function, the normalization factor.

We have thus generalized the concepts of Gibbs function and chemical potentials in thermochemistry to non-normalized statistical counting number in terms of the entropy function $S(\{m_i\})$. We have discovered that the Gibbs–Kirkwood potential functions are the conjugate variables $\{u_i\}$. Last but not least, by substituting $u_i = \partial S/\partial m_i$ into the EoS in Eq. (22), one obtains a Hamilton–Jacobi equation for the homogeneous entropy function $S(\{m_i\})$:[4,13]

$$\sum_{i=1}^{n} p_i \exp\left(-\frac{\partial S}{\partial m_i} \right) - 1 = 0, \tag{25}$$

as an alternative to the Gibbs–Duhem equation in Eq. (23).

For a finite state space with independent and identically distributed (i.i.d.) statistical data, the results in Eqs. (22)–(25) provide a complete portrayal of its "macroscopic statistical thermodynamics". Extending this formalism to continuous state spaces is outside the scope of the present work.

3.2. Markov chain large deviations and conjugate process

Another direction for a generalization of the above results is on correlated statistical data: Beyond the i.i.d. empirical counting, most repeated measurements of a stationary stochastic signal have a temporal correlation which can be represented by a Markov process $X_t, t = 0, 1, \ldots$, with transition probability matrix $\{p_{ij}\}$ between consecutive measurements. The LDRF of the counting frequencies along a path is then[9]

$$I(\{\nu_i\}) = -\inf_{\text{all } \mu_i > 0} \sum_{i=1}^{n} \nu_i \log \left(\frac{1}{\mu_i} \sum_{j=1}^{n} \mu_j p_{ji} \right). \tag{26}$$

One sees that if $p_{ij} = \pi_j$ that is independent of i, then Eq. (26) becomes

$$-\inf_{\text{all } \mu_i > 0} \sum_{i=1}^{n} \nu_i \log \left(\frac{\pi_i}{\mu_i} \sum_{j=1}^{n} \mu_j \right) = \sum_{i=1}^{n} \nu_i \log \frac{\nu_i}{\pi_i}.$$

The infimum in Eq. (26) can be solved from a system of nonlinear equations for the μ_i's for each given set of $\{\nu_i\}$:

$$\sum_{i=1}^{n} \nu_i \left(\frac{\mu_k p_{ki}}{\sum_{j=1}^{n} \mu_j p_{ji}} \right) = \nu_k, \ 1 \le k \le n. \tag{27a}$$

$$I(\{\nu_i\}) = \sum_{i=1}^{n} \nu_i \log \left(\frac{\nu_i}{w_i} \right), \quad w_i = \nu_i \left(\sum_{j=1}^{n} \frac{\mu_j p_{ji}}{\mu_i} \right)^{-1}, \tag{27b}$$

where $\{w_i\}$ denotes a set of normalization, which is determined for the given set of $\{\nu_i\}$. Equation (27) has two very different interpretations: First in data statistical terms, the observed counting frequency $\{\nu_i\}$ is the stationary distribution of a modified Markov chain with transition probability $\hat{p}_{ij}$:

$$\begin{aligned}
\hat{p}_{ij} &= \frac{\mu_j p_{ji}}{\sum_{k=1}^{n} \mu_k p_{ki}} = \frac{(\mu_j/\sum_{l=1}^{n} \mu_l)\, p_{ji}}{\sum_{k=1}^{n} \mu_k p_{ki}/\sum_{l=1}^{n} \mu_l} \\
&= \frac{\Pr\{X_t = j, X_{t+1} = i\}}{\Pr\{X_{t+1} = i\}},
\end{aligned} \tag{28}$$

which is the Bayesian posterior distribution for X_t given $X_{t+1} = i$, under the original Markov process with $\Pr\{X_t = j, X_{t+1} = i\} = (\mu_j/\sum_{l=1}^{n} \mu_l)\, p_{ji}$. Equation (27a) finds the hidden $\{\mu_i\}$ as a change of measure[21] based on the observed $\{\nu_i\}$. Equation (27b) also has a thermodynamic interpretation: $\log(\nu_i/w_i)$ in Eq. (27b) is a Gibbs–Kirkwood potential representing the deviation of observed statistical frequency $\{\nu_i\}$ from the stationary probability $\{\pi_i\}$ dictated by the $\{p_{ij}\}$. It is zero when $\nu_i = \pi_i\ \forall i$, which yields $\mu_k = \pi_k$ and $\nu_i = w_i$ from Eq. (27). In terms of ν's and w's, $\hat{p}_{ij} =$

$(\mu_j p_{ji}/\mu_i)w_i/\nu_i$. The transition probability p_{ij} in standard Markov theory actually is based on an implicit assumption that each "state" counts as 1; there is certainly no intrinsic reason for this assumption and the $\{\mu_i\}$ is a "correction" that accounts for observed stationary empirical counting frequency. In other words, from observations $\{\nu_i\}$ that are in disagreement with $\{\pi_i\}$, the first-order correction is not on the assumption for the transition probability $\{p_{ij}\}$; rather one modifies the naive, uniform counting measure used for each and every state i: from 1 to μ_i.

From a mathematical modeling standpoint, finding $\{p_{jk}\}$ for an underlying Markov system and finding a conjugate process $\{\hat{p}_{jk}\}$ for the empirical frequency of measurements on the stochastic system are equivalent. Any given set of observed counting frequencies $\{\nu_i\}$ is represented as the stationary distribution of the conjugate Markov process w.r.t. the hidden measure $\{\mu_i\}$ that serves as a tilting,[22] and represents the measurement on probability. Empirical counting of the occurrences of a discrete state via an observable is not directly related to the probability of the state; it is related to the ratio of its probability measure and a "Lebesgue measure of the state" defined by the observable.

For a continuous-time Markov chain with $p_{ij} = (\mathrm{e}^{\boldsymbol{Q}\Delta t})_{ij}$ where $\boldsymbol{Q} = \{q_{ij}\}$ is a transition rate matrix, one has Eq. (26) in the limit of $\Delta t \to 0$:

$$-\frac{1}{\Delta t}\inf_{\text{all }\mu_i>0}\sum_{i=1}^{n}\nu_i\log\left(\frac{1}{\mu_i}\sum_{j=1}^{n}\mu_j\left(\delta_{ji}+q_{ji}\Delta t\right)\right)$$
$$\to -\inf_{\text{all }\mu_i>0}\sum_{i,j=1}^{n}\frac{\mu_j q_{ji}\nu_i}{\mu_i}. \tag{29}$$

The conjugate process w.r.t $\{\mu_i\}$ has transition probability rate matrix

$$\hat{q}_{ij} = \frac{\mu_j q_{ji}}{\mu_i} \text{ if } i \neq j,\ \hat{q}_{ii} = -\sum_{j\neq i}^{n}\hat{q}_{ij} = -\sum_{j\neq i}^{n}\frac{\mu_j q_{ji}}{\mu_i}, \tag{30}$$

and $\{\nu_i\}$ is its stationary measure. Actually $\hat{q}_{ij} = \mu_j q_{ji}/\mu_i - r_i\delta_{ij}$ where $r_i = q_{ii} - \hat{q}_{ii}$ has another interpretation: the instantaneous rate of change in the logarithm of the non-normalized measure μ_i of state i following the Markov process: $\mathrm{d}\log\mu_i/\mathrm{d}t = \sum_{k=1}^{n}\mu_k q_{ki}/\mu_i = r_i$. When $\forall i\ \mu_i = \pi_i$, $r_i = 0$. The larger a $|r_i|$ value for a state i, the faster the state probability relaxes to π_i. In this context, the Gibbs–Kirkwood potential for a Markov chain also acquires a kinetic meaning.

3.3. Continuous-time Markov chain with detailed balance

Interestingly the infimum in Eq. (29) can be explicitly solved if $\{q_{ij}\}$ has detailed balance, $\pi_i q_{ij} = \pi_j q_{ji}$, where $\{\pi_i\}$ is the stationary distribution of the continuous-time Markov chain. One has

$$\sum_{i,j=1}^{n}\frac{\mu_j q_{ji}\nu_i}{\mu_i} \geq \sum_{i=1}^{n}q_{ii}\nu_i + 2\sum_{i>j}^{n}\left(q_{ji}\nu_i q_{ij}\nu_j\right)^{\frac{1}{2}}, \tag{31}$$

where the equality holds true when $\mu_j q_{ji}\nu_i/\mu_i = \mu_i q_{ij}\nu_j/\mu_j$. That is, if $(\mu_j/\mu_i)^2 = \nu_j\pi_j/(\nu_i\pi_i)\ \forall i, j$. Therefore the entropy function has a closed form[23]

$$\frac{I(\{\nu_i\})}{\Delta t} = \sum_{i,j=1}^{n}\sqrt{\nu_i}\,\hat{\Pi}_{ij}\sqrt{\nu_j}, \tag{32}$$

in which matrix $\hat{\boldsymbol{\Pi}}$ is semi-positive definite with elements $\hat{\Pi}_{ij} = -\sqrt{\pi_i}q_{ij}/\sqrt{\pi_j} = \hat{\Pi}_{ji}$. As a LDRF, $I(\{\nu_i\}) \geq 0$ and the equality holds true when $\nu_i = \pi_i$. This corresponds to matrix $\{q_{ij}\}$ having an eigenvalue 0 with eigenvector $(1,\ldots,1)$. The Legendre–Fenchel transform of Eq. (32) yields an eigenvalue problem that solves the $\{\nu_i\}$ for given $u_i = \partial I/\partial\nu_i$, $(1 \leq i \leq n)$:

$$\sum_{k=1}^{n}\left[\,\hat{\Pi}_{ik} - (u_i+\lambda)\delta_{ik}\right]\sqrt{\nu_k} = 0, \tag{33}$$

in which λ again represents the gauge freedom that arises from the constraint among ν's. The λ is the principal eigenvalue of matrix $\left[\hat{\boldsymbol{\Pi}} - \mathrm{diag}(u_1,\ldots,u_n)\right]$, with the corresponding eigenvector being $\{\sqrt{\nu_i}\}$. As the LFT of $I(\{\nu_i\})$ in Eq. (32), $\lambda(u_1,\ldots,u_n)$ is a convex function.[24]

Equation (32) offers a novel narrative for the statistical behavior of correlated empirical counting frequencies from a time-reversible Markov process: The empirical counting frequencies are split into $\sqrt{\nu_i}$ and $\sqrt{\nu_j}$, they are captured by a symmetric matrix $\sqrt{\nu_i\nu_j}$ whose trace is 1. This resembles von Neumann's density operator. Equation (33) further suggests a novel, symmetric eigenvalue problem for $\sqrt{\nu_i}$. The physical significance of this set of results remains to be explored.

4. Discussion

4.1. The nature of Gibbs' theory

We identify Gibbs' theories of statistical mechanics and chemical thermodynamics put together as a program that is very different from Boltzmann's original objective of developing a mechanical theory of heat, e.g., thermodynamic principles as mathematical consequences of classical mechanics. Gibbs' theory, with certain justifications, puts the notion of temperature which represents stochastic mechanical motions as an *a priori* concept and canonical ensemble theory as a first principle. This theory reproduces all the known relationships in classical equilibrium thermodynamics of macroscopic systems by taking the thermodynamic limit. One of the key results in this regard is the Gibbs–Duhem equation, which we now recognize as the differential form of the thermodynamic equation of state. Gibbs' theory however is also applicable to "small systems". The entirety of statistical chemistry and soft matter physics, including theories of non-ideal solutions, polymers, polyelectrolytes, proteins, and the great successes these applications have brought about, is a testimony to the validity

of the scientific theory, which is far beyond Boltzmann's original vision, if one can surmise. Boltzmann's theory has a goal; Gibbs' is open ended.

The macroscopic limit gives rise to theories of emergent phenomena, such as phase transitions. This insight originated in Yang and Lee's work in 1952,[25] who departed from the then traditional approach of McMillan–Mayer cluster expansion. By 1972, this had become one of the fundamental tenets for emergence.[10,26]

The current work provides the Gibbsian theory a new foundation and its limitation. It is a mathematical consequence within the theory of probability, not due to the large size thermodynamic limit, but rather due to the law of large numbers of measurements whose empirical counting measure converges to probability. In other words, the "law" articulated in the Gibbsian theory is due to universal statistical behavior of frequency as an empirical measurement for probability, which is a mathematical rather than physical notion.

4.2. Maximum entropy principle re-interpreted

Maximum entropy principle (MEP)[27] and its dynamic counterpart maximum calibre (MaxCal)[28] have a great number of followers in engineering and data science. The theory of large deviations offers the very statistical origin for such a principle: The key is to distinguish the abstract mathematical concept of "probability measure" associated with $(\Omega, \mathcal{F}, \mathbb{P})$ and the statistical counting frequency that is widely used as the surrogate for the former. The latter can be experimentally measured. Very recently, K. A. Dill has called for a distinction between entropy as a scalar quantity of a system and entropy as a "function of the normalized counting frequencies".[29]

4.3. Gauge freedom, normalization and free energy of ground state

The Legendre–Fenchel dual to the Shannon–Sanov entropy of counting frequency $\{\nu_i\}$ naturally gives rise to the notion of Gibbs–Kirkwood potential function $\{u_i\}$ together with a gauge freedom λ which is mathematically related to the normalization of the m's. It defines through a gauge fixing the "ground" state in which its Helmholtz free energy λ is zero. It is important to recognize that strictly speaking a set of definitive $\{\nu_i\}$ only results from infinite many measurements $M \to \infty$: Given a set of $\{\nu_i\}$ or alternatively given a set of $\{u_i\}$, one defines a statistical or "thermodynamic" system with permanence, certainty, without randomness. In classical physics, such a system is defined mechanically through its energy function. In the present work, the mechanics is replaced by the underlying probability and an entropy function from LDT.

4.4. Bottom up, top down, and bottom down

In traditional physics, theories for explaining and predicting natures are often termed "fundamental" and/or "phenomenological", some philosophers of science also invented terms such as "top down" and "bottom up".[30] The Gibbsian theory as understood from the present work provides a surprisingly penetrating analytic device for carrying out the vision of R. B. Laughlin who coined a third term " bottom down".[31] In the past, the theory at the fundamental level has always been mechanics, which provides the notion of mechanical energy and mechanical work, from which thermodynamics arises. But as many practitioners of biophysical chemistry have recognized,[32] the relationship between the macromolecular Boltzmann weights (or Kirkwood's potential of mean force) one has at hand and the constitutive mechanical Hamiltonian of point masses is strenuous even at its best. The present work, in agreement with Ref. [1], again suggests that the traditional view might not be the necessary path to thermodynamic behavior. Statistical law could even provide an alternative origin for the very notion of internal energy, via the Gibbs–Kirkwood potential function.

The eigenvalue equation in Eq. (33) exhibits several intriguing features: First, the square of its eigenvector, $\sqrt{\nu_k}$, yields the "probability" $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)$. This L_2 interpretation is in sharp contrast to the eigenvalue problem in standard Markov theory which is in L_1. Second, the symmetric matrix $\hat{\boldsymbol{\Pi}}$ is known as the discrete counterpart of the Laplace operator in continuous space, and u_i is analogous to the potential energy function in the Schrödinger operator. This uncanny resemblance between quantum mechanical Schrödinger equation and Eq. (33) that arises from a statistical analysis parallel to the Gibbsian thermodynamics for counting deserves serious further investigation. Last but not least, Eq. (32) is another entropy function which bears no resemblance to the Shannon entropy in terms of logarithmic probability. It has expanded the repertoire for the concept of entropy. Further investigations into the counting frequency of correlated events might reveal more unexpected behavior.[33]

Acknowledgement

We thank R. H. Austin, J. C. Baez, W. R. Cannon, K. A. Dill, C. Jia, L. F. Thompson and H. Zhao for helpful discussions.

References

[1] Smith E 2020 *Entropy* **22** 1137
[2] Qian H 2022 *J. Chem. Theory Comput.* **18** 6421
[3] Qian H 2024 *Entropy* **26** 1091
[4] Ge H and Qian H 2016 *Phys. Rev. E* **94** 052150
[5] Ge H and Qian H 2017 *J. Stat. Phys.* **166** 190
[6] Hill T L 1963 *Thermodynamics of Small Systems* (New York: Dover)
[7] Lu Z and Qian H 2022 *Phys. Rev. Lett.* **128** 150603

[8] Oono Y 1989 *Progr. Theor. Phys. Supp.* **99** 165
[9] Dembo A and Zeitouni O 1998 *Large Deviations Techniques and Applications*, 2nd Ed. (New York: Springer)
[10] Chibbaro S, Rondoni L and Vulpiani A 2014 *Reductionism, Emergence and Levels of Reality* (New York: Springer)
[11] Angelini E and Qian H 2023 *J. Phys. Chem. B* **127** 2552
[12] Rockafellar R T 1996 *Convex Analysis* (NJ: Princeton Univ. Press)
[13] Miao B, Qian H and Wu Y S 2024 arXiv:2406.02405
[14] Shannon C E and Weaver W 1949 *The Mathematical Theory of Communication* (Champaign: Univ. Ill. Press)
[15] Hobson A 1969 *J. Stat. Phys.* **1** 383
[16] Shore J E and Johnson R W 1980 *IEEE Trans. Inf. Th.* **26** 26
[17] Guggenheim E A 1933 *Modern Thermodynamics by the Methods of Willard Gibbs* (New York: Methuen & Co.)
[18] Callen H B 1991 *Thermodynamics and an Introduction to Thermostatistics*, 2nd Ed. (New York: John Wiley & Sons)
[19] Huang K 1963 *Statistical Mechanics* (New York: John Wiley & Sons)
[20] Kirkwood J G 1935 *J. Chem. Phys.* **3** 300
[21] Yang Y J and Qian H 2020 *Phys. Rev. E* **101** 022129
[22] Barato A C and Chetrite R 2015 *J. Stat. Phys.* **160** 1154
[23] Baldi P and Piccioni M 1999 *Stat. Probab. Lett.* **41** 107
[24] Cohen J 1981 *Proc. Amer. Math. Soc.* **81** 657
[25] Yang C N and Lee T D 1952 *Phys. Rev.* **87** 404
[26] Anderson P W 1972 *Science* **177** 393
[27] Jaynes E T 2003 *Probability Theory: The Logic of Science* (Cambridge: Cambridge Univ. Press)
[28] Ghosh K, Dixit P D, Agozzino L and Dill K A 2020 *Annu. Rev. Phys. Chem.* **71** 213
[29] Dill K A 2021 *The maximum caliber principle for modeling stochastic dynamic processes*
[30] Salmon W C 1990 *Four Decades of Scientific Explanation* (Pittsburgh: Univ. Pittsburgh Press)
[31] Laughlin R B 2006 *A Different Universe: Reinventing Physics from the Bottom Down* (New York: Basic Books)
[32] Schellman J A 1997 *Biophys. Chem.* **64** 7
[33] Hoffmann B 2011 *The Strange Story of the Quantum* (New York: Dover)