A New Formalism of the Boltzmann-Gibbs Statistical Mechanics

Xiang Gao

(Dated: June 30, 2020)

The Boltzmann-Gibbs (BG) statistical mechanics (SM) has two widely accepted formalisms: BG formalizes SM as a physical theory, while E.T. Jaynes’ considers it as a statistical theory. Though being developed for more than a century, both approaches today still have unresolved philosophical questions. Such as, for the BG formalism: Why can we identify the information entropy with thermodynamic entropy? Why can we identify physical quantities with ensemble averages? The Jaynes’ maximum entropy approach is also criticized for valid only when the observer’s knowledge is in a particular form with a particular prior. This article proposes a way to formalize statistical mechanics as a statistical theory while avoiding its maximum entropy principle. By applying a less controversial subset of assumptions taken from the BG formalism, surprisingly, we can prove that the generalized Boltzmann distribution is the only distribution mathematically consistent with the Legendre structure of thermodynamics. By formalizing SM this way, we avoid these unresolved issues.

I. INTRODUCTION

The celebrated Boltzmann-Gibbs (BG) statistical mechanics is the core of modern physics in describing matters and radiations. This theory dates back to Boltzmann[1] and Gibbs[2]. At the beginning of its invention, such theory contained assumptions coming from nowhere, such as the unit of volume like $h^3 N$, and had problems describing physical phenomena such as the specific heat of diatomic gas and the Gibbs paradox. While these problems were later cleared up by quantum mechanics, some other issues still do not have an utterly satisfying answer even today[3].

One of such problem is: why can we describe a physical system by considering its imaginary duplications, a.k.a. “ensemble”? This question has two subquestions: (Q1) why can we identify the thermodynamic entropy with the information entropy? (i.e., why is the Boltzmann entropy formula correct? Note that, this identity appears as a postulate in the BG formalism); and (Q2), why can we identify physical quantities with the ensemble average?

For question Q1, the literature contains not only studies supporting the idea that the information is physical but also studies criticizing it. Examples of supporting research include Landauer’s principle[4–9] and Szilard’s engine[10–13]. However, these examples only show that information is physical on specific circumstances, and it by no means indicates we could quite generally identifying the thermodynamic entropies with information entropy. Criticizes about this idea are straightforward: unlike $dV$ or $dE$, which has clear definitions by geometry or mechanics, $dS$ only makes sense through $dQ$ and $T$; however, there are no concepts similar to temperature or heat in information theory[14–18].

For question Q2, the most famous answer is through ergodicity: Macroscopic measurements take much longer time than microscopic processes. Therefore, the results obtained from measurements are always the time average for an infinite time. With infinite time, the system’s trajectory will visit all microstates; therefore, that time average equals an average over microstates. With the development of measure theory, this ergodicity was proved to be wrong[19–21], but fortunately, later development of the ergodic theory saved the identity between time average and ensemble average by discarding Boltzmann’s original interpretation[22–24]. Subtle issues on the foundation of the ergodic theory are still being criticized, and readers are referred to the review[3] for more information. A bigger problem of this answer is, this approach only converts the question from “why can ensemble average describe physical quantities” to “why can time average describe physical quantities,” and the answer to the latter question is not clear either. As pointed out in[25, 26], “measurement needs time” should not be confused with “measuring the time average.” Besides, averaging on a large (no matter how large it is) finite time may result in very different conclusions from averaging on infinite time. For example, if we can only measure infinite time behavior, why can we even see non-equilibrium phenomena, considering that all systems will reach equilibrium give large enough time?

In 1957, E.T. Jaynes[22, 23] introduced the concept of “subjective statistical mechanics,” where he suggests considering the statistical mechanics as a statistical theory instead of as a physical theory. The key to this formalism is the maximum entropy principle: with limited information on the system, the most unbiased probability distribution is the one that maximizes the Shannon entropy under given constraints, where for the case of constraint $\sum_\omega \Pr(\omega) \cdot E(\omega) = \text{constant}$, the resulting distribution would be the Boltzmann distribution. Doing so makes statistical mechanics free from any physical assumptions such as ergodicity: “whether or not the results agree with experiment, they still represent the best estimates that could have been made on the basis of the information available.” Although this idea was accepted by most physicists, there are also sharp criticisms about this theory. Criticizers claim that this theory is valid.
only if the observer has a specific form of knowledge and correctly chosen prior. As a result, in practice, the accuracy of this theory must be empirically validated. Therefore the theory “lacks a solid foundation” and “has not led to any new concrete result.”

This article borrows the idea from E.T. Jaynes for considering statistical mechanics as a statistical theory. That is, we use an ensemble with a probability distribution to describe the system of interest, but we do not care about the physical meaning of this distribution, e.g., it does not matter whether or not this distribution is the distribution of a system in equilibrium with a reservoir. By applying a less controversial subset of assumptions taken from the BG formalism, surprisingly, we can prove that the generalized Boltzmann distribution is the only distribution mathematically consistent with the Legendre structure of thermodynamics. By establishing a statistical theory without using the maximum entropy principle, we take its benefits while avoiding its problems. By using fewer assumptions compared to the BG formalism, this formalism avoids Q1 and answers Q2.

Before moving to the main result of this article, the author would like to refer readers interested in foundations of statistical mechanics to reviews in 1955, 1979, and a more recent survey.

II. ASSUMPTIONS

Our formalism uses the following assumptions:

**Assumption 1.** A physical system can be described by an ensemble of imaginary duplicates of that system, each duplicate has its state called microstate, denoted by \( \omega \), and each microstate is associated with weight \( f(\omega) \). The probability density function \( \Pr(\omega) \) can be computed by normalizing weights of microstates, that is

\[
\Pr(\omega) = \frac{f(\omega)}{\sum_{\omega'} f(\omega')}
\]

Although it is not apparent, the above assumption is also implicitly assumed in the BG formalism of statistical mechanics. In the BG formalism, we study a system in contact with a reservoir. The interaction of the contact is assumed to be weak in the sense that the set of microstates of the system and the reservoir is the cartesian product of the set of microstates of the system of interest and the of the reservoir. This weakly interacting assumption implies that the probability density of the system of interest is proportional to the number of microstates in the reservoir, which is mathematically identical to the assumption above.

**Assumption 2.** Thermodynamic state functions are described by the ensemble average of random variables.

The above assumption is the assumption that connects the microscopic world with the macroscopic world in the BG formalism. Only with this assumption, the BG formalism can write down the famous \( F = k_B T \log Z \) and derive the rest state functions. For the BG formalism, this assumption is questioned as Q2. Since we are considering the statistical mechanics as a statistical theory, in our construct, this assumption is not claiming that the system’s physical quantities equal averages over the distribution when the system is in a particular imaginary circumstance. Instead, this assumption should be interpreted as: we are looking for a probability distribution that can describe physical quantities as averages. It would be possible that people can find a statistical theory that describes physical quantities in a different functional form. We expect that different functional forms lead to different statistics, such as the Tsallis statistics.

**Assumption 3.** In general, lower energy states are more preferred than higher energy states. At high temperatures, the gap in the probability density between lower and higher energy states is smaller than at low temperatures. At the limit of infinite temperature, all microstates have the same probability density.

In the BG formalism, the above assumption is a common sense and does not need to be explicitly stated. However, in our formalism, we are not assuming any physical meaning in our distribution. So this has to be stated as an assumption to filter out potentially non-physical results.

It worth mentioning that, different from the BG formalism who assumes the Boltzmann entropy formula, we do not make any assumption on entropy, which allows us to be free from the problem of Q1.

III. THE THEORY

Consider a thermodynamic system with generalized forces \( X_1, \ldots, X_n \), \( Y_1, \ldots, Y_m \) and generalized coordinates \( \chi_1, \ldots, \chi_n \), \( y_1, \ldots, y_m \). The first law of thermodynamics of that system reads

\[
dU = TdS + \sum_{\eta=1}^n X_\eta d\chi_\eta + \sum_{\eta=1}^m Y_\eta dy_\eta
\]

We want to study an ensemble parametrized by \((T, X_1, \ldots, X_n, y_1, \ldots, y_m)\). In this setup, \( E, x_1, \ldots, x_n \) are random variables and \( Y_1, \ldots, Y_m \) are statistical quantities of that ensemble. For a microstate \( \omega \), we denote the value of random variables at \( \omega \) as \( E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)} \). We use different letters \( X \) vs \( Y \) to distinguish ensemble parameters from statistical quantities. Also, we use \( \chi \) vs \( x \) to distinguish thermodynamic state functions from random variables. The key of our formalism is the following theorem:

**Theorem 1.** An ensemble that describes a thermodynamic system whose first law reads as equation and
who satisfies assumptions § and § and satisfies the generalized Boltzmann distribution:

$$\Pr(\omega) \propto \exp \left[ \sum_{n=1}^{\infty} \frac{X_n x_n^{(\omega)}}{k_B T} - \frac{E^{(\omega)}}{k_B T} \right]$$  (3)

**Proof.** This proof uses some lemmas, which are stated and proved in section § From assumption § we can write $\Pr(\omega)$ as

$$\Pr(\omega) \propto f\left(E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}; T, X_1, \ldots, X_n\right)$$  (4)

Let $\beta = \frac{1}{k_B T}$, $\tilde{X}_n = \beta X_n$ and $\tilde{Y}_n = \beta Y_n$. Instead of writing $f$ as a function of $(T, X_1, \ldots, X_n)$, we will write it as a function of $(\beta, \tilde{X}_1, \ldots, \tilde{X}_n)$:

$$\Pr(\omega) \propto f\left(E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}; \beta, \tilde{X}_1, \ldots, \tilde{X}_n\right)$$  (5)

Rewrite the first law of thermodynamics equation § with $\beta, \tilde{X}_1, \ldots, \tilde{X}_n$, we get

$$\frac{dS}{k_B} = \beta dU - \sum_{\eta=1}^{n} \tilde{X}_\eta d\chi_\eta - \sum_{\eta=1}^{n} \tilde{Y}_\eta d\eta_\eta$$  (6)

do Legendre transformation to get a state function $B$ with natural variables $\beta, \tilde{X}_1, \ldots, \tilde{X}_n, y_1, \ldots, y_m$, we have

$$B = \frac{S}{k_B} - \beta U + \sum_{\eta=1}^{n} \tilde{X}_\eta \chi_\eta$$  (7)

$$dB = -U d\beta + \sum_{\eta=1}^{n} \chi_\eta d\tilde{X}_\eta - \sum_{\eta=1}^{n} \tilde{Y}_\eta d\eta_\eta$$  (8)

therefore

$$U = \left<E^{(\omega)}\right> = -\frac{\partial B}{\partial \beta}$$  (9)

$$\chi_\eta = \left<x^{(\omega)}_\eta\right> = \frac{\partial B}{\partial X_\eta}$$  (10)

The normalization constant (partition function) for equation § is

$$Z = \sum_\omega f_\omega$$  (11)

where $f_\omega$ is short for

$$f\left(E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}; \beta, \tilde{X}_1, \ldots, \tilde{X}_n\right)$$  (12)

Then equation § and equation § becomes

$$\sum_\omega \frac{E^{(\omega)} f_\omega}{Z} = -\frac{\partial B}{\partial \beta}$$  (13)

From basic multivariable calculus, we have $\frac{\partial^2 B}{\partial X_\eta \partial \beta}$. Therefore

$$\frac{\partial}{\partial X_\eta} \sum_\omega \frac{E^{(\omega)} f_\omega}{Z} + \frac{\partial}{\partial \beta} \sum_\omega \frac{x^{(\omega)}_\eta f_\omega}{Z} = 0$$  (15)

which simplifies to

$$\sum_\omega \left[ E^{(\omega)} \frac{\partial (f_\omega/Z)}{\partial X_\eta} + x^{(\omega)}_\eta \frac{\partial (f_\omega/Z)}{\partial \beta} \right] = 0$$  (16)

the above equality should always be true, regardless of the details of the system and microstates, the only way to guarantee this is to have

$$E^{(\omega)} \frac{\partial (f_\omega/Z)}{\partial X_\eta} + x^{(\omega)}_\eta \frac{\partial (f_\omega/Z)}{\partial \beta} = 0$$  (17)

for all $\omega$s. Apply the same thing to $\frac{\partial^2 B}{\partial \sigma \partial X_\eta} = \frac{\partial^2 B}{\partial X_\eta \partial \sigma}$, and from lemma § we know that $f$ must have the form

$$g\left(\zeta, E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}\right)$$

where

$$\zeta = \beta E^{(\omega)} - \sum_{\eta=1}^{n} \tilde{X}_\eta x^{(\omega)}_\eta$$  (18)

Let $G$ be an antiderivative of $g$ with respect to $\zeta$, that is,

$$G' = g\left(\zeta, E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}\right)$$  (19)

We use the prime $'$ exclusively for derivative with respect to the first argument $\zeta$ while keeping other arguments $E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}$ constant. Let $K = \sum_\omega G\left(\zeta; E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}\right)$, it is easy to show that

$$\frac{\partial K}{\partial \beta} = \sum_\omega E^{(\omega)} g_\omega = Z \left<E^{(\omega)}\right> = -Z \cdot \frac{\partial B}{\partial \beta}$$  (20)

$$\frac{\partial K}{\partial X_\eta} = -\sum_\omega x^{(\omega)}_\eta g_\omega = -Z \cdot \left<x^{(\omega)}_\eta\right> = -Z \cdot \frac{\partial B}{\partial X_\eta}$$  (21)

where $g_\omega$ is short for

$$g\left(\beta E^{(\omega)} - \sum_{\eta=1}^{n} \tilde{X}_\eta x^{(\omega)}_\eta, E^{(\omega)}, x_1^{(\omega)}, \ldots, x_n^{(\omega)}\right)$$  (22)

Note that $K, Z, B$ all have the same set of natural variables $\beta, \tilde{X}_1, \ldots, \tilde{X}_n, y_1, \ldots, y_m$, so equation § and equation § can be condensed as

$$dK = -Z \cdot dB$$  (23)
Properties of exact differential tell us that $K$, $Z$, and $B$ must have function relationship between each other. $K$ and $Z$ are both functionals with parameters $\beta, \tilde{X}_1, \ldots, \tilde{X}_n$ that map functions of $\omega$ (random variables $E(\omega), x_1(\omega), \ldots, x_n(\omega)$) to numbers. If the random variables change by a small amount $\delta E(\omega), \delta x_1(\omega), \ldots, \delta x_n(\omega)$, then the change of these functionals are:

$$\delta K = \sum_{\omega} \left[ \left( \frac{\partial G}{\partial E(\omega)} + \beta g_\omega \right) \delta E(\omega) + \sum_\eta \left( \frac{\partial g}{\partial x_\eta} - \tilde{X}_\eta g_\omega \right) \delta x_\eta(\omega) \right]$$

$$\delta Z = \sum_{\omega} \left[ \left( \frac{\partial g}{\partial E(\omega)} + \beta g'_\omega \right) \delta E(\omega) + \sum_\eta \left( \frac{\partial g}{\partial x_\eta} - \tilde{X}_\eta g'_\omega \right) \delta x_\eta(\omega) \right]$$

where the $\frac{\partial}{\partial E(\omega)}$ and $\frac{\partial}{\partial x_\eta}$ are partial derivatives keeping $\zeta$ constant:

$$\frac{\partial}{\partial E(\omega)} \bigg|_{\zeta, x_1(\omega), \ldots, x_n(\omega)}$$

$$\frac{\partial}{\partial x_\eta(\omega)} \bigg|_{\zeta, E(\omega), x_1(\omega), \ldots, x_n(\omega), \tilde{X}_\eta(\omega), \ldots, \tilde{X}_n(\omega)}$$

The function relationship between $K$ and $Z$ requires $\delta K = C \left( \beta, \tilde{X}_1, \ldots, \tilde{X}_n \right) \delta Z$ to be true for all $\delta E(\omega), \delta x_1(\omega), \ldots, \delta x_n(\omega)$, where $C \left( \beta, \tilde{X}_1, \ldots, \tilde{X}_n \right)$ is some constant that must not depend on $E(\omega), x_1(\omega), \ldots, x_n(\omega)$ but could depend on $\beta, \tilde{X}_1, \ldots, \tilde{X}_n$. Then

$$\frac{\partial G}{\partial E(\omega)} + \beta g_\omega = C \left( \beta, \tilde{X}_1, \ldots, \tilde{X}_n \right) \left[ \frac{\partial g}{\partial E(\omega)} + \beta g'_\omega \right]$$

$$\frac{\partial G}{\partial x_\eta} - \tilde{X}_\eta g_\omega = C \left( \beta, \tilde{X}_1, \ldots, \tilde{X}_n \right) \left[ \frac{\partial g}{\partial x_\eta} - \tilde{X}_\eta g'_\omega \right]$$

From lemma $\delta K = C \left( \beta, \tilde{X}_1, \ldots, \tilde{X}_n \right)$ is a constant that does not depend on $\beta, \tilde{X}_1, \ldots, \tilde{X}_n$. Denote it as $C_1$. Define

$$\hat{L} = \frac{\partial}{\partial E(\omega)} + \beta \frac{\partial}{\partial \zeta}$$

then equation $28$ can be written as $\hat{L} G = C_1 \hat{L} g$. Since $\hat{L}$ is a linear operator, we have $\hat{L} (G - C_1 g) = 0$. The kernel of $\hat{L}$ contains functions of the form $\varphi \left( \zeta - \beta E(\omega) \right)$. Do the same thing to equation $29$ and summarize, we see that $G - C_1 g$ must have the form

$$G - C_1 g = \varphi \left( \zeta - \beta E(\omega) + \sum_{\eta=1}^n \tilde{X}_\eta x_\eta(\omega) \right)$$

However, $\zeta - \beta E(\omega) + \sum_{\eta=1}^n \tilde{X}_\eta x_\eta(\omega) = 0$, we then have $G - C_1 g = C_2$ where $C_2$ is another constant. Taking derivative of both side, we get $g = C_1 \hat{L} g'$, which immediately leads to

$$g \left( \zeta, E(\omega), x_1(\omega), \ldots, x_n(\omega) \right) =$$

$$C_2 \left( E(\omega), x_1(\omega), \ldots, x_n(\omega) \right) \cdot \exp \left( \zeta / C_1 \right)$$

and

$$G \left( \zeta, E(\omega), x_1(\omega), \ldots, x_n(\omega) \right) =$$

$$C_1 \cdot C_2 \left( E(\omega), x_1(\omega), \ldots, x_n(\omega) \right) \cdot \exp \left( \zeta / C_1 \right)$$

where $C_2 \left( E(\omega), x_1(\omega), \ldots, x_n(\omega) \right)$ is some constant that must not depend on $\beta, \tilde{X}_1, \ldots, \tilde{X}_n$ but could depend on $E(\omega), x_1(\omega), \ldots, x_n(\omega)$. From assumption $\delta$ we must have $C_1 < 0$. Since $C_1$ is a constant multiplied towards the temperature, from lemma $\delta$ we can choose $C_1 = -1$ without loss of generality. Defining

$$\Lambda(\omega) = C_2 \left( E(\omega), x_1(\omega), \ldots, x_n(\omega) \right)$$

we have

$$\Pr(\omega) \propto \Lambda(\omega) \cdot \exp \left[ \sum_{\eta=1}^n \frac{X_\eta x_\eta(\omega)}{k_B T} - \frac{E(\omega)}{k_B T} \right]$$

At $T \to \infty$, the probability density function $\Pr(\omega) \to \Lambda(\omega)$. From assumption $\delta$ $\Lambda(\omega)$ must be a constant. This concludes the proof.

The procedure to obtain all other thermodynamic state functions is the same as in the BG formalism: In the proof of theorem $1$ it is easy to observe that $K = -Z$. Then from equation $23$ we know that $B = \log Z + C_3$. Since $B$ is extensive, $C_3$ must vanish. We then have $B = \log Z$. Let $J = k_B T \cdot B = k_B T \log Z$ and substitute into equation $7$ and equation $8$ we get

$$J = TS - U + \sum_{\eta=1}^n X_\eta \chi_\eta$$

$$dJ = SdT + \sum_{\eta=1}^n \chi_\eta dX_\eta - \sum_{\eta=1}^m Y_\eta d\eta$$

we can then obtain all state functions following the standard procedure.
IV. LEMMAS AND THEIR PROOF

Lemma 1. For a function of 4 variables \( f(a, b, c, d) \), if

\[
\frac{\partial f}{\partial b} \bigg|_{a \ b \ c \ d} + c \frac{\partial f}{\partial d} \bigg|_{a \ b \ c \ d} = 0
\]

then there exists a function \( g \) such that \( f(a, b, c, d) = g(ad - bc, a, c) \).

Proof. Let us call \((a, b, c, d)\) the old coordinates, and define a new coordinates \((u, v, w, x)\)

\[
\begin{align*}
u &= a \\
v &= c \\
w &= ad - bc \\
x &= ad + bc
\end{align*}
\]

then the reverse transformation is

\[
\begin{align*}
a &= u \\
c &= v \\
d &= w + z \\
b &= \frac{zu}{v}
\end{align*}
\]

Evaluating partial derivatives in new coordinate, we have

\[
\left. \frac{\partial f}{\partial x} \right|_{uvw} = \frac{\partial f}{\partial a} \left|_{a \ b \ c \ d} \cdot \frac{1}{2a} + \frac{\partial f}{\partial b} \left|_{a \ b \ c \ d} \cdot \frac{1}{2c} \right.
\]

\[
= \frac{1}{2ac} \left( c \frac{\partial f}{\partial d} \bigg|_{abc} \bigg| + a \frac{\partial f}{\partial d} \bigg|_{abc} \right)
\]

that \( f \) does not depend on \( x \). Therefore, it is a function of only \( u, v, w \).

Lemma 2. Let \( f, g \) and \( C \) be functions, and \( x, y, a, b \) be variables, then \( f(ax + by, x, y) = C(a, b) \cdot g(ax + by, x, y) \) implies that \( C(a, b) \) is a constant that does not depend on \( a, b \).

Proof. For fixed \( x, y \), the set of all possible values of \((a, b)\) that have \( ax + by = z \), where \( z \) is a constant, is a line. When \( x, y \) and \( ax + by \) are all fixed, the values of \( f \) or \( g \) does not change, so that line \( C(a, b) \) must also be a constant. This is true for all values of \( x, y \) and \( z \), that is, on all possible lines, \( C(a, b) \) is a constant. Since different lines cross, \( C(a, b) \) then must be a constant that does not depend on \( a, b \).

Lemma 3. If we scale the temperature by \( \frac{1}{\alpha} \) and the entropy by \( \alpha \), we don’t change any physics.

Proof. Let’s begin our proof by reviewing how the theory of equilibrium statistical mechanics is built. The procedure starts from defining the entropy of the microcanonical ensemble as \( S = k_B \log \Omega \), and establish a system in thermal equilibrium with a reservoir who defines \( T \). The number of microstates of the reservoir \( \Omega_r \) is therefore given by

\[
\Omega_r = \exp \left( \frac{S_r}{k_B} \right)
\]

taking the power series of \( S_r \) at \( E_{total} \) with respect to the energy of the system \( E_s \) and keeping up to the first order, we get

\[
S_r (E_{total} - E_S) = S_r (E_{total}) - \frac{\partial S_r}{\partial E_r} \cdot E_S
\]

from the first law of thermodynamics, we know that

\[
\frac{\partial S_r}{\partial E_r} = \frac{1}{T}
\]

combining equation 42, equation 43 and equation 44, we get the Boltzmann distribution

\[
\Omega_r = C \cdot \exp \left( -\frac{E_s}{k_B T} \right)
\]

where \( C = \exp (S_r (E_{total})) \) is a constant that does not depend on \( E_S \). In the above procedure, the temperature scale is introduced by defining entropy as \( S = k_B \log \Omega \). The constant \( k_B \) in that equation gets propagated along the logic chain and determines the temperature scale together with the first law of thermodynamics.

If we instead started by defining \( S' = a k_B \log \Omega \), following the same logic we will get

\[
\Omega_r = \exp \left( \frac{S'_r}{k_B a T'} \right)
\]

in this case, the first law of thermodynamics will give

\[
\frac{\partial S'_r}{\partial E_r} = \frac{1}{T'}
\]

where \( T' \) is the temperature in the new scale. The Boltzmann distribution in the new scale will look like

\[
\Omega_r = C \cdot \exp \left( -\frac{E_s}{k_B a T'} \right)
\]

Temperature scales are artificial, but probabilities are physical. So equation 48 must match with equation 45.

To prove that they match, we use the first law of thermodynamics to find the relationship between \( T \) and \( T' \):

\[
\begin{align*}
dU &= T dS - pdV \\
dU &= T' dS' - pdV \Rightarrow T' &= \frac{T}{\alpha}
\end{align*}
\]

substitute equation 49 into equation 48 we get an exact match with equation 45 which concludes this proof.

V. CONCLUSION

We just showed: We want to find a probability distribution over microstates that can be used to describe a physical system. We do not care how to establish a physical environment to implement this probability distribution on the interested system. For mathematical
simplicity, we want our distribution in the form that each microstate is assigned a weight (assumption), and the physical quantities described by averages (assumption). After doing some math, we see that the generalized Boltzmann distribution is the only distribution that satisfies all the requirements.

In comparison, the BG formalism: We want to study a physical system in contact and equilibrium with a reservoir. In such a setup, the distribution is in a form that each microstate is assigned a weight. If we further assume the information entropy and thermodynamic quantities are described by averages. Therefore, we do not know how to connect this distribution to thermodynamics yet, so we further assume thermodynamic quantities are described by averages. With this assumption, the theory is completed.

Compared with E.T. Jaynes's formalism, although we both consider statistical mechanics as a statistical theory, the formalism in this article does not use anything about maximum entropy estimates. Therefore the criticisms about E.T. Jaynes's formalism does not apply to the formalism proposed here. Compared with BG's formalism, we do not assume anything about entropy; therefore, the issue is successfully avoided.

Our formalism also answers Q2: It is not true that physical quantities must be described by averages. It has nothing to do with ergodicity. It is possible to describe physical quantities with a different functional, but each functional has its corresponding probability density. If averages are selected, then this distribution coincides with the distribution of the system when in contact and equilibrium with a reservoir. If other functional is chosen, a different distribution will be needed. For example, it is possible to study a physical system utilizing the distribution representing the scenario when that system is in equilibrium with another finite-sized system. In such a configuration, the physical quantities have to be represented by something different from averages.

One justification of the above answer to Q2 is if we replace assumption with as in and is easy to verify that with slight modifications our proof to theorem still holds, and we will obtain the Tsallis statistics instead of the BG statistics. The Tsallis statistics are widely used in describing complex systems. Interested readers are referred to for more details on Tsallis Statistics and its applications. The website maintained by Constantino Tsallis is also a good source for study.

VI. ACKNOWLEDGEMENT

The author thanks Constantino Tsallis for the discussion on the connection with the Tsallis statistics.

[1] L. Boltzmann, *Wissenschaftliche Abhandlungen* edited by F. Hasenohrl (Cambridge University Press, Cambridge, 2012).
[2] J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, 1902) p. 207.
[3] R. Frigg, in *The Ashgate Companion to Contemporary Philosophy* edited by D. Rickles (2011) Chap. 3: A Field Guide to Recent Work on the Foundations of Statistical Mechanics.
[4] R. Landauer, IBM Journal of Research and Development (1961).
[5] J. A. Vaccaro and S. M. Barnett, in *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* (2011) arXiv:1004.5330.
[6] A. Béret, A. Arakelyan, A. Petrosyan, S. De Liberato, and M. Ueda, Physical Review Letters (2011), 10.1103/PhysRevLett.106.070401.
[7] Y. Jun, M. Gavrilov, and J. Bechhoefer, Physical Review Letters (2014), 10.1103/PhysRevLett.113.190601.
[8] J. Hong, B. Lambson, S. Dhuey, and J. Bokor, Science Advances (2016), 10.1126/sciadv.1501492.
[9] R. Gaudenzi, E. Burzuri, S. Maegawa, H. S. van der Zant, and F. Luis, “Quantum landauer erasure with a molecular nanomagnet,” (2018).
[10] L. Szilard, Zeitschrift für Physik (1929), 10.1007/BF01341284.
[11] J. J. Park, K. H. Kim, T. Sagawa, and S. W. Kim, Physical Review Letters (2013), 10.1103/PhysRevLett.111.230402.
It is natural to start with a simple form, if, after try and error, we find that there is no such form of distribution that can lead to useful theory, we can try something more complicated.

Hint: the conclusion of theorem 1 needs to be replaced with the $q$-distribution, and the equation 19 needs to be replaced with $G' = g^q$.