Article

Equivalence of Partition Functions Leads to Classification of Entropies and Means

Michel S. Elnaggar [1],[†] and Achim Kempf [2],*

[1] Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada;
E-Mail: mselnaggar@ieee.org

[2] Department of Applied Mathematics, University of Waterloo, 200 University Avenue West,
Waterloo, Ontario N2L 3G1, Canada

[†] Present address: Bell Mobility Inc., Mississauga, Ontario L4W 5N2, Canada.

* Author to whom correspondence should be addressed; E-Mail: akempf@uwaterloo.ca;
Tel.: +1-519-888-4567 (ext. 35462); Fax: +1-519-746-4319.

Abstract: We derive a two-parameter family of generalized entropies, S_{pq}, and means m_{pq}. To this end, assume that we want to calculate an entropy and a mean for n non-negative real numbers $\{x_1,...,x_n\}$. For comparison, we consider $\{m_1,...,m_k\}$ where $m_i = m$ for all $i = 1,...,k$ and where m and k are chosen such that the l^p and l^q norms of $\{x_1,...,x_n\}$ and $\{m_1,...,m_k\}$ coincide. We formally allow k to be real. Then, we define k, $\log k$, and m to be a generalized cardinality k_{pq}, a generalized entropy S_{pq}, and a generalized mean m_{pq} respectively. We show that this family of entropies includes the Shannon and Rényi entropies and that the family of generalized means includes the power means (such as arithmetic, harmonic, geometric, root-mean-square, maximum, and minimum) as well as novel means of Shannon-like and Rényi-like forms. A thermodynamic interpretation arises from the fact that the l^p norm is closely related to the partition function at inverse temperature $\beta = p$. Namely, two systems possess the same generalized entropy and generalized mean energy if and only if their partition functions agree at two temperatures, which is also equivalent to the condition that their Helmholtz free energies agree at these two temperatures.

Keywords: cardinality; dimensionality; entropy; equivalence; free energy; information and thermodynamics; ℓ^p norm; mean; partition function; Shannon and Rényi axioms

PACS Codes: 05.70.-a, 02.50.-r

---

1. Introduction

Two of the most basic concepts of thermodynamics are: (a) the average of measurement outcomes and (b) the uncertainty or entropy about measurement outcomes. Consider, for example, a physical system, *A*, that is in contact with a heat bath at some fixed temperature, *i.e.*, a canonical ensemble. The measurement of the system's energy can return any one of its energy eigenstates. What then is (a) the mean energy to expect and (b) how uncertain is the prediction of the measured energy state?

We notice that, in principle, many different notions of energy mean and many different measures of entropy could be employed here. Of course, in thermodynamics, the Boltzmann factor weighted *mean* as well as the Shannon/von Neumann *entropy* are of foremost importance. In this paper, we show that also other important notions of average such as the harmonic mean, the geometric mean and the arithmetic mean arise naturally, along with generalized notions of entropy including Rényi entropies [1], all unified in a two-parameter family of notions of means and notions of entropies.

To this end, consider systems (canonical ensembles) in a heat bath. We begin by considering the simplest kind of system, namely the type of system which possesses only one energy level, *E*. Let us denote its degeneracy by *k*. Unambiguously, we should assign that system the mean $m := E$ and the entropy $S := \log k$. Let us denote these simple one-level systems by the term *reference system*.

Now, let *X* be a system with arbitrary discrete energy levels. Our aim is to assign *X* a mean and an entropy by finding that reference system *M* which is in some sense equivalent to *X*. Then we assign *X* the same value for the mean and entropy as the reference system *M*.

But how do we decide if a reference system is in some sense equivalent to system *X*? Given that we want the reference system *M* and system *X* to share two properties, namely a mean and an entropy, we expect any such condition for the equivalence of two systems to require two equations to be fulfilled. Further, since the properties of systems are encoded in their partition function $Z(X)$, we expect that these two equations can be expressed in terms of the partition functions of the two systems in question.

To this end, let us adopt what may be considered the simplest definition. We choose two temperatures, T_1 and T_2 and we define that a reference system is (T_1, T_2)-equivalent to system *X* if the partition functions of the two systems coincide with each other at these two temperatures. Since the Helmholtz free energy A_H obeys $A_H = -K_B T \log(Z)$, where K_B is the Boltzmann constant, this is the same as saying that two systems are put in the same equivalence class if their Helmholtz free energies coincide at these two temperatures.

This allows us now to assign any system *X* a mean and an entropy. We simply find its unique (T_1 and T_2)-equivalent reference system *M*. Then the mean and entropy of *X* are defined to be the mean and the entropy of the reference system *M*.

Clearly, the so-defined mean and entropy of a system *X* now actually depend on *two* temperatures, namely (T_1, T_2). As we will show below, in the limit when we let the two temperatures become the same temperature, we recover the usual Boltzmann factor-weighted mean, *i.e.*, the usual mean energy, along with the usual Shannon/von Neumann entropy.

For general (T_1, T_2), we cover more, however. Namely, we naturally obtain a unifying 2-parameter family of notions of mean that includes for example the geometric, the harmonic, the arithmetic and the root-mean-square (RMS) means. And we obtain a unifying 2-parameter family of notions of entropy that, for example, includes the Rényi family of entropies.

To be precise, let us assume that a system X has only discrete energy levels, $\{E_i\}$, where i enumerates all the energy levels counting also possible degeneracies. Notice that $\{E_i\}$ is formally what is called a multiset, because its members are allowed to occur more than once. Similarly, let us also collect the exponentiation of the negative energies $x_i := \exp(-E_i)$ in the multiset $\{x_i\}$. Either multiset can be used to describe the same thermodynamic system X. Let $\beta := 1/K_B T$ denote the inverse temperature, where K_B is the Boltzmann constant. The partition function of system X, *i.e.*, the sum of its Boltzmann factors, then reads:

$$Z_\beta(X) = \sum_i e^{-\beta E_i} = \sum_i x_i^{\beta} \tag{1}$$

For later reference, note that the partition function is therefore related to the l^p norm of $X = \{x_1, ..., x_n\}$ for $p = \beta$ through $Z_\beta(X) = \| X \|_\beta^\beta$.

Now the key definition is that we call two physical systems (β_1, β_2)-equivalent if their partition functions coincide at the two inverse temperatures (β_1, β_2), *i.e.*, systems X and M are (β_1, β_2)-equivalent if $Z_{\beta_1}(X) = Z_{\beta_1}(M)$ and $Z_{\beta_2}(X) = Z_{\beta_2}(M)$. To be more explicit, one may also call such systems (β_1, β_2)-partition function equivalent, or also (β_1, β_2)-Helmholtz free energy equivalent, but we will here use the term (β_1, β_2)-equivalent for short.

In particular, for any given system X, let us consider the (β_1, β_2)-equivalent reference system M which possesses just one energy level, with energy E_0 and degeneracy k, where we formally allow k to be any positive number. E_0 and k are then determined by the two conditions that the partition function of M is to coincide with that of X at the two inverse temperatures β_1 and β_2. Then, we define $S_{\beta_1, \beta_2}(X) := \log k$ to be the generalized entropy, and $m_{\beta_1 \beta_2}(X) := E_0$ to be the generalized mean energy of system X with respect to the temperatures (β_1, β_2).

We will explore the properties of these families of generalized entropies and means in the subsequent sections. First, however, let us consider the special limiting case when the two temperatures coincide (*i.e.*, $\beta_1 = \beta_2 = \beta$). As will be detailed in the subsequent sections of the manuscript, in this limiting case, the two equivalence conditions of partition functions can be shown to reduce to:

$$\left(Z_\beta(X) \right)^{1/\beta} = \left(Z_\beta(M) \right)^{1/\beta} \tag{2a}$$

$$\frac{\partial}{\partial \beta} \left(Z_\beta(X) \right)^{1/\beta} = \frac{\partial}{\partial \beta} \left(Z_\beta(M) \right)^{1/\beta} \tag{2b}$$

which can be shown to be equivalent to:

$$Z_\beta(X) = Z_\beta(M) \tag{3a}$$

$$\sum_i E_i \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}} = E_0 \tag{3b}$$

The conditions (3a) and (3b) physically mean that systems X and M have the same *partition function* and *average energy*, respectively, at the inverse temperature β. Notice that this is also the same as saying that the two systems have the same average *energy* and the same Helmholtz free *energy* at the inverse temperature β. Now by employing either pair of conditions, (2) or (3), we then recover indeed the usual thermodynamic entropy of the system X, which is given in the Shannon form at the inverse temperature β by

$$S_{\beta\beta}(X) = -\sum_i \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}} \log\left(\frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}}\right) \tag{4}$$

The proofs of Equations (2)–(4) are straightforward. We will spell them out in detail in the subsequent sections where the setting is abstract mathematical.

Before we begin the mathematical treatment, let us remark that entropy is not only a cornerstone of thermodynamics but it is also crucial in information theory. Due to its universal significance, measures of uncertainty in the form of an entropy have been proposed by physicists and mathematicians for over a century [2]. Our approach here for deriving a generalized family of entropies was originally motivated by basic questions regarding the effective dimensionality of multi-antenna systems (e.g., [3–5]). After initial attempts in [6] and later in [5,7], we here give for the first time a comprehensive derivation with the proofs and we also include the family of generalized means.

The manuscript is organized as follows. In Section 2, we introduce the proposed family of entropies and means mathematically, and also show some special cases thereof. The axiomatic formulation is presented in Section 3, followed by the study of resulting properties in Section 4. Proofs are provided in the appendices.

2. Mathematical Definition of Generalized Entropies and Means

Let $X = \{x_i : i = 1, 2, \ldots, K_X\}$ be a multiset of real non-negative numbers, where K_X denotes the cardinality of X. We assume that X possesses at least one non-zero element. Further, let p, q be arbitrary fixed real numbers obeying $pq \geq 0$. Let $M = \{m_i : m_i = m, m > 0, i = 1, 2, \ldots, k\}$ be a reference multiset possessing exactly one real positive element m, which is of multiplicity $k \geq 1$. We introduce the following definitions:

- N_X is the number of non-zero elements of X, therefore $N_X \leq K_X$. To simplify notation, the subscript X is omitted when dealing with one multiset at hand.
- $K_{X\max}$ is the multiplicity of the maximum elements of X
- $K_{X\min}$ is the multiplicity of the minimum elements of X

Our objective is to determine suitable values for m and k, possibly non-integer, that can serve as mean and effective cardinality of X, respectively, namely by imposing a suitable criterion for the equivalence of X to a reference multiset M. Having two unknowns (m and k) in M, we need two equivalence conditions. We choose to impose the equivalence of the *p*-norms and the *q*-norms:

$$\begin{aligned}
\|X\|_p &\equiv \|M\|_p = mk^{1/p} \\
\|X\|_q &\equiv \|M\|_q = mk^{1/q}
\end{aligned} \tag{5}$$

Here, the *p*-norm $\|\cdot\|_p$ is defined as usual through:

$$\|X\|_p := \left(\sum_{i=1}^{K_X} (x_i)^p \right)^{1/p} \tag{6}$$

with the proviso that $(x_i)^p$ is replaced by 0 if $x_i = 0$ and $p \le 0$. We remark that, for $p < 1$, (6) is merely a quasi-norm since the triangle inequality does not hold. Note the singularity $\lim_{p \to 0^+} \|X\|_p = \infty$.

Solving for k and m in (5), we obtain:

$$k_{pq}(X) = \left[\|X\|_p / \|X\|_q \right]^{1 / \left(\frac{1}{p} - \frac{1}{q} \right)} \qquad p \ne q$$

$$m_{pq}(X) = \left[\|X\|_p^p / \|X\|_q^q \right]^{\frac{1}{p-q}} \qquad p \ne q \tag{7}$$

We call $k_{pq}(X)$ and $m_{pq}(X)$ the norm-induced *effective cardinality* and *generic mean* of order p, q for the multiset *X*, respectively. Let us now express (7) in a logarithmic form and define the entropy $S_{pq}(X)$ as follows:

$$S_{pq}(X) := \log k_{pq}(X) = \frac{\log \|X\|_p - \log \|X\|_q}{\frac{1}{p} - \frac{1}{q}} \qquad p \ne q \tag{8a}$$

$$\log m_{pq}(X) = \frac{p \log \|X\|_p - q \log \|X\|_q}{p - q} \qquad p \ne q \tag{8b}$$

Notice that $S_{pq}(X) = S_{qp}(X)$ and $m_{pq}(X) = m_{qp}(X)$, *i.e.*, both the entropy and mean are symmetrical with respect to the order p, q.

Next, we express $S_{pq}(X)$ and $m_{pq}(X)$ in the limiting case when $p \to q$. For $S_{qq}(X)$ we find:

$$\begin{aligned} S_{qq}(X) &:= \log k_{qq}(X) \\ &= \lim_{p \to q} S_{pq}(X) \\ &= \lim_{p \to q} \frac{\log \|X\|_p - \log \|X\|_q}{\frac{1}{p} - \frac{1}{q}} \\ &\overset{\text{l'Hôpital}}{=} -q^2 \frac{\frac{\partial}{\partial q} \|X\|_q}{\|X\|_q} \\ &= -\sum_{i=1}^{K} \frac{x_i^q}{\sum_{j=1}^{K} x_j^q} \log \left(\frac{x_i^q}{\sum_{j=1}^{K} x_j^q} \right) \end{aligned} \tag{9}$$

where the last step is obtained by straightforward manipulations. Similarly, we find for $m_{qq}(X)$:

$$m_{qq}(X) = \lim_{p \to q} m_{pq}(X) = \|X\|_q \Big/ k_{qq}^{1/q}$$

$$\log m_{qq}(X) = \lim_{p \to q} \frac{p \log \|X\|_p - q \log \|X\|_q}{p - q}$$

$$\overset{\text{l'Hôpital}}{=} \log \|X\|_q + q \frac{\dfrac{\partial}{\partial q} \|X\|_q}{\|X\|_q} \tag{10}$$

$$= \frac{\dfrac{\partial}{\partial q} \|X\|_q^q}{\|X\|_q^q}$$

$$= \sum_{i=1}^{K} \frac{x_i^q}{\sum_{j=1}^{K} x_j^q} \log x_i$$

where in the second to last step, we used the fact that $\dfrac{d}{dx} g(x)^{f(x)} = g^f \left\{ \dfrac{df}{dx} \log g + \dfrac{f}{g} \dfrac{dg}{dx} \right\}$, and the last

step is obtained by straightforward manipulations.

It is worthwhile to mention the following useful relation linking $S_{qq}(X)$, $\|X\|_q$ and $m_{qq}(X)$, which

is readily deduced from (9) and (10):

$$\log m_{qq}(X) = \log \|X\|_q - S_{qq}(X)/q \tag{11}$$

We remark that in the early phase of this work [5], each author independently suggested either $k_{pq}(X)$ or $k_{qq}(X)$ as two possible *distinct* notions for the *effective cardinality*. In [7], it was reported that the *average* energy and *Shannon entropy* of a thermodynamic system are obtained by starting from equivalence of partition functions of two systems at two temperatures when the two temperatures coincide as mentioned in the introduction. Clearly, the limiting operation in (9) makes the connection and establishes (7) as the general definition of this norm-induced family of entropies and means.

In fact, for the case of degenerate order, ($p \to q$), the quantities $k_{qq}(X)$, $S_{qq}(X)$, and $m_{qq}(X)$, could have been obtained as well through a differential equivalence of the q-norm. To see this, we impose the following two conditions:

$$\|X\|_q \equiv \|M\|_q$$
$$\frac{\partial}{\partial q} \|X\|_q \equiv \frac{\partial}{\partial q} \|M\|_q \tag{12}$$

After employing $\dfrac{d}{dx} g(x)^{f(x)} = g^f \left\{ \dfrac{df}{dx} \log g + \dfrac{f}{g} \dfrac{dg}{dx} \right\}$ and solving for k and m, we ultimately obtain

$k_{qq}(X)$ and $m_{qq}(X)$ as given by (9) and (10). The condition (12) is the mathematical equivalent of the aforementioned physical condition (2) imposed on the two thermodynamic systems, which yielded the Shannon entropy form (4).

From (9), it is obvious that $S_{qq}(X)$ is the Shannon entropy of the distribution $\left\{ x_i^q \Big/ \sum_{i=1}^{K} x_i^q \right\}$, which

is called the escort distribution of order q [8] of $\left\{ x_i \Big/ \sum_{i=1}^{K} x_i \right\}$. On the other hand, $S_{pq}(X)$ is a more

general expression of the Rényi entropy of order α. For a probability distribution $P = \{p_i\}$, the Rényi entropy of order α is given by [9]:

$$S_\alpha^{(R)}(P) := \frac{1}{1-\alpha} \log \frac{\sum_i p_i^{\;\alpha}}{\sum_i p_i}, \quad \alpha > 0, \alpha \neq 1 \tag{13}$$

By setting the order $p = 1$ in S_{pq}, we obtain from (8):

$$S_{1q}(X) = \frac{\log \sum_{i=1}^K x_i^{\;q} - q \log \sum_{j=1}^K x_j}{1-q} \quad q > 0, q \neq 1$$

$$= \frac{1}{1-q} \log \sum_{i=1}^K \left(\frac{x_i}{\sum_{j=1}^K x_j} \right)^q \tag{14}$$

By comparing (13) and (14), we readily identify $S_{1q}(X)$ as the Rényi entropy of order q for a complete statistical distribution given by $\{x_i / \sum_{i=1}^K x_i\}$, where the multiset elements add to 1. Formally:

$$S_{1q}(X) = S_q^{(R)}\left(\{x_i / \sum_{i=1}^K x_i\} \right) \tag{15}$$

In the degenerate case (when $q \to 1$), $S_{11}(X)$ is the Shannon entropy of the latter distribution. For $p \neq 1$, $S_{pq}(X)$ from (8) can be rearranged as a generalization of (13):

$$S_{pq}(X) = \frac{p}{p-q} \log \sum_{i=1}^K \left(\frac{x_i^{\;p}}{\sum_{j=1}^K x_j^{\;p}} \right)^{q/p}$$

$$= \frac{1}{1-q/p} \log \sum_{i=1}^K \left(\frac{x_i^{\;p}}{\sum_{j=1}^K x_j^{\;p}} \right)^{q/p} \tag{16}$$

$$= S_{q/p}^{(R)}\left(\{x_i^{\;p} / \sum_{i=1}^K x_i^{\;p}\} \right)$$

which can be viewed as the Rényi entropy of order q/p for the pth order escort distribution $\{x_i^{\;p} / \sum_{i=1}^K x_i^{\;p}\}$.

Rényi defined his entropy for $p, q \geq 0$. We relax this condition further and allow $S_{pq}(X)$ and $m_{pq}(X)$ to be defined for any real indices p, q such that $pq \geq 0$. Accordingly, we obtain the following properties:

$$S_{-p,-q}(\{x_i\}) = S_{p,q}(\{x_i^{-1} : x_i \neq 0\})$$

$$m_{-p,-q}(\{x_i\}) = m_{p,q}^{-1}(\{x_i^{-1} : x_i \neq 0\}) \tag{17}$$

When at least one order p, q is zero, we find the interesting results:

$$k_{0q}(X) = N_X \tag{18}$$

$$S_{0q}(X) = \log N_X \tag{19}$$

$$m_{0q}(X) = \begin{cases} \left(\dfrac{\sum_{i=1}^{N_X} x_i^{\,q}}{N_X} \right)^{1/q} & x_i \neq 0, q \neq 0 \\[4mm] \left(\prod_{i=1}^{N_X} x_i \right)^{1/N_X} & x_i \neq 0, q \to 0 \end{cases} \tag{20}$$

We recognize $S_{0q}(X)$ in (19) as the Hartley entropy [10], which will be shown later to be the maximum value of any entropy. From (20), we obtain a famous family of generic *p*-means of the non-zero elements of *X*: particularly $m_{0,-\infty}, m_{0,-1}, m_{0,1}, m_{0,2}, m_{0,\infty}$ are the minimum, harmonic mean, arithmetic mean, root-mean-square mean, and maximum, respectively. In the limiting case $q \to 0$, we obtain $m_{0,0}$, which is the geometric mean. In Table 1, we summarize these and other particular cases of means and entropies at specific p, q. The key point is that each p, q uniquely defines an entropy S_{pq} with a corresponding mean m_{pq}, such that each pair of S_{pq} and m_{pq} is coupled in this sense.
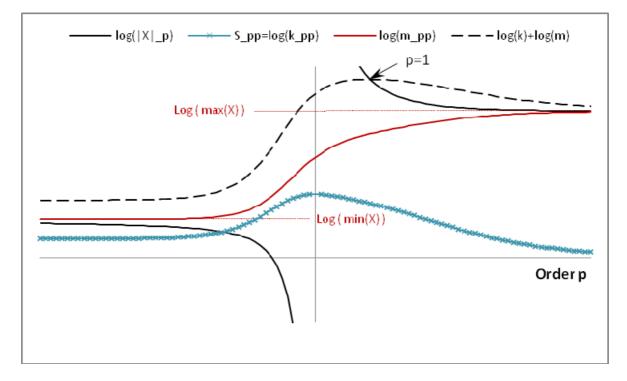
Table 1. Special cases of S_{pq} and m_{pq}. Note that $S_{pq} = S_{qp}$ and $m_{pq} = m_{qp}$ (Property 4.3).

Order p, q	$S_{pq}(X)$	S_{pq} name	$m_{pq}(X)$	m_{pq} name
$0, q \; q \neq 0$	$\log N_X$	Boltzmann-Hartley entropy	$\left(\dfrac{\sum_{i=1}^{N_X} x_i^{\,q}}{N_X} \right)^{1/q}$	Generic q mean. Specific q values are harmonic (-1), arithmetic (1), root-mean-square (2), maximum (∞), minimum ($-\infty$)
$0, 0$	$\log N_X$	Boltzmann-Hartley entropy	$\left(\prod_{i=1}^{N_X} x_i \right)^{1/N_X}$	Geometric mean
∞, ∞	$\log K_{X\max}$		$\max(X)$	maximum
$\infty, q \; q > 0$	$\log\left(K_{X\max} + \displaystyle\sum_{x_i \neq \max(x_i)} \left(\dfrac{x_i}{\max(x_i)} \right)^q \right)$		$\max(X)$	maximum
$-\infty, -\infty$	$\log K_{X\min}$		$\min(X)$	minimum
$-\infty, q \; q < 0$	$\log\left(K_{X\min} + \displaystyle\sum_{x_i \neq \min(x_i)} \left(\dfrac{x_i}{\min(x_i)} \right)^q \right)$		$\min(X)$	minimum
$1, q \; q > 0$	$\dfrac{1}{1-q} \log \displaystyle\sum_{i=1}^{K} \left(\dfrac{x_i}{\sum_{j=1}^{K} x_j} \right)^q$	Rényi entropy, order q, of the complete distribution $\{x_i / \sum_{i=1}^{K} x_i\}$	$\dfrac{\sum_{i=1}^{K} x_i}{\exp S_{1q}(X)}$	"Rényi-like" mean
$1, 1$	$-\displaystyle\sum_{i=1}^{K} \dfrac{x_i}{\sum_{j=1}^{K} x_j} \log\left(\dfrac{x_i}{\sum_{j=1}^{K} x_j} \right)$	Gibbs-Shannon entropy	$\dfrac{\sum_{i=1}^{K} x_i}{\exp S_{11}(X)}$	"Shannon-like" mean

A typical plot for $\|X\|_p$, S_{pp} and m_{pp} is shown in log scale in Figure 1, illustrating some of the properties to be discussed hereafter. In particular, we notice:

- $\log\|X\|_p$ has a two-sided singularity at $p=0$.

- S_{pp} is non-decreasing/non-increasing for negative/positive p, respectively, and is guaranteed to be maximized at $p=0$. This is discussed more generally in Property 4.6.

- m_{pp} ranges from $\min(X)$ to $\max(X)$ and is always non-decreasing with respect to p. This is discussed more generally in Property 4.6.

- $p=1$ has a specific property of making $m_{pp}k_{pp}=\|X\|_p$.

Figure 1. Typical plot for $\|X\|_p$, S_{pp} and m_{pp} in log scale.

3. An Axiomatic Approach to the Generalized Entropies and Means

In order to simplify the conceptual underpinnings, let us now describe the generalized entropies S_{pq} and means m_{pq}, through two simple axioms.

3.1. Axioms for the Generalized Entropy

Let p,q be fixed real numbers obeying $pq \geq 0$. Consider a map, S_{pq}, which maps multisets of positive real numbers into the real numbers. We call S_{pq} a generalized entropy of order p,q if it obeys the following two axioms:

Entropy Axiom 1:

$S_{pq}(M)$ of a uniform multiset $M=\{m_0,m_0,\ldots,m_0\}$, $m_0>0$, with multiplicity k, equals $\log k$, (where the base of the logarithm is arbitrarily chosen), *i.e.*,

$$M = \{m_i : m_i = m_0, m_0 > 0, i = 1, 2, \ldots, k\} \Rightarrow S_{pq}(M) = \log k \tag{21}$$

Entropy Axiom 2:

If $p \neq q$, the map S_{pq} depends only on the ratio of the multiset's p and q norms, *i.e.*, S_{pq} is some function, f_{pq}, of this ratio:

$$S_{pq}(X) := f_{pq}\left(\frac{\|X\|_p}{\|X\|_q}\right) \qquad p \neq q \qquad (22a)$$

If $p = q$, the map S_{qq} depends only on the ratio of the multiset's q norm to its derivative, *i.e.*, S_{pq} is some function, f_q, of this ratio:

$$S_{qq}(X) := f_q\left(\frac{\partial/\partial q \|X\|_q}{\|X\|_q}\right) \qquad p \to q \qquad (22b)$$

To see that (22b) arises in the limit from (22a) we notice that, since the logarithm is strictly monotone, (22a) is equivalent to saying that S_{pq} is some function h_{pq} of some finite number times the logarithm of the ratio of norms:

$$\begin{aligned}
S_{pq}(X) &= h_{pq}\left(\frac{1}{p-q}\log\frac{\|X\|_p}{\|X\|_q}\right) \\
&= h_{pq}\left(\frac{\log\|X\|_p - \log\|X\|_q}{p-q}\right)
\end{aligned} \qquad (23)$$

Choosing $p = q + \varepsilon$ and taking the limit $\varepsilon \to 0$ we obtain (22b) with $f_q = h_{qq}$.

Proposition: The two entropy axioms in (21) and (22) uniquely define $S_{pq}(X)$ and $S_{qq}(X)$, namely as given in section 2 by (8) and (9).

Proof:

Entropy Axiom 2 implies that for any two multisets X and Y:

$$\frac{\|X\|_p}{\|X\|_q} = \frac{\|Y\|_p}{\|Y\|_q} \qquad \Rightarrow S_{pq}(X) = S_{pq}(Y) \qquad for \ \ p \neq q \qquad (24a)$$

$$\frac{\partial/\partial q \|X\|_q}{\|X\|_q} = \frac{\partial/\partial q \|Y\|_q}{\|Y\|_q} \qquad \Rightarrow S_{qq}(X) = S_{qq}(Y) \qquad for \ \ p = q \qquad (24b)$$

Choosing for Y the uniform multiset M of Axiom 1, and taking the logarithm of both sides yields:

$$\begin{aligned}
\log\frac{\|X\|_p}{\|X\|_q} &= \left(\frac{1}{p} - \frac{1}{q}\right)\log k \qquad p \neq q \\
\frac{\partial/\partial q \|X\|_q}{\|X\|_q} &= -\frac{\log k}{q^2} \qquad\quad p = q
\end{aligned} \qquad (25)$$

Using that $S_{pq}(X) = S_{pq}(Y) = S_{pq}(M) = \log k$, we now uniquely obtain the formulas for $S_{pq}(X)$ and $S_{qq}(X)$ given in (8) and (9), respectively. We note that the functions f_{pq} and f_q are therefore:

$$f_{pq}(\bullet) = \frac{\log(\bullet)}{\dfrac{1}{p} - \dfrac{1}{q}} \qquad p \neq q$$

$$f_q(\bullet) = -q^2(\bullet) \qquad p = q \tag{26}$$

Remarks:

- Even though k is treated as an integer representing the multiplicity in Axiom 1, this condition is tacitly relaxed in Axiom 2 to include non-integer values, which we may call the effective cardinality (or effective dimensionality) of order p, q.

- The logarithmic measure of Axiom 1 is directly connected to the celebrated Boltzmann entropy formula $S^{(B)} = K_B \log W$, where K_B is the Boltzmann constant and W is the number of the microstates in the system. The logarithmic measure is also connected to the so-called "Hartley's measure" [11,12], which indicates the non-specificity [13] and does not require a probability distribution assumption. In Axiom 1, a multiset of equal positive numbers is all that is required. In fact, Axiom 1 encompasses the additivity and monotonicity axioms [12,13], which are equivalent to the earlier Khinchin's axioms of additivity, maximality and expansibility [14].
- In Axiom 2, note that the p-norm definition is relaxed to include the values $p < 1$, which would result in the triangle inequality to be violated should the multisets be treated as vectors.

3.2. Axioms for the Generalized Mean

We define the pth moment of the multiset $X = \{x_i\}$ as:

$$p^{th} \, moment(X) := \|X\|_p^p = \sum_i x_i^p \tag{27}$$

The nomenclature "p^{th} moment" is motivated by the fact that for the density function $r(x) := \sum_i \delta(x - x_i)$, where $\delta(x)$ is the Dirac delta function, the pth moment is indeed:

$$
\begin{aligned}
p^{th} \, moment(X) &:= \int x^p r(x) dx \\
&= \int x^p \sum_i \delta(x - x_i) dx \\
&= \sum_i x_i^p = \|X\|_p^p
\end{aligned}
\tag{28}
$$

Let p, q be fixed real numbers obeying $pq \geq 0$. Consider a map, m_{pq}, which maps multisets of positive real numbers into the real numbers. We call m_{pq} a generalized mean of order p, q if it obeys the following two axioms:

Mean Axiom 1:

$m_{pq}(M)$ of a uniform multiset $M = \{m_0, m_0, \ldots, m_0\}$, $m_0 > 0$ is m_0. Formally:

$$M = \{m_i : m_i = m_0, m_0 > 0, i = 1, 2, \ldots, k\} \Rightarrow m(M) := m_0 \tag{29}$$

Mean Axiom 2:

If $p \neq q$, the map m_{pq} depends for any multiset X only on the ratio of the multiset's p^{th} and q^{th} moments, $\| X \|_p^p$ and $\| X \|_q^q$, *i.e.*, m_{pq} is some function, g_{pq}, of their ratio:

$$m_{pq}(X) := g_{pq}\left(\frac{\| X \|_p^p}{\| X \|_q^q} \right) \qquad p \neq q \tag{30a}$$

If $p = q$, the map m_{qq} is a function only of the ratio:

$$m_{qq}(X) := g\left(\frac{\partial / \partial q \, \| X \|_q^q}{\| X \|_q^q} \right) \qquad p = q \tag{30b}$$

The fact that (30b) is the limit of (30a) follows by the same reasoning as in (23).

Proposition: The two mean axioms in (29) and (30) uniquely define m_{pq}, namely as given in (7) and (10).

Proof: Axiom 2 implies for any two multisets X and Y that:

$$\frac{\| X \|_p^p}{\| X \|_q^q} = \frac{\| Y \|_p^p}{\| Y \|_q^q} \qquad \Rightarrow m_{pq}(X) := m_{pq}(Y) \qquad p \neq q$$

$$\frac{\partial / \partial q \, \| X \|_q^q}{\| X \|_q^q} = \frac{\partial / \partial q \, \| Y \|_q^q}{\| Y \|_q^q} \qquad \Rightarrow m_{qq}(X) := m_{qq}(Y) \qquad p = q \tag{31}$$

Choosing the multiset Y to be the uniform multiset M from (29) we obtain

$$\frac{\| X \|_p^p}{\| X \|_q^q} = m_0^{p-q} \qquad p \neq q$$

$$\frac{\partial / \partial q \, \| X \|_q^q}{\| X \|_q^q} = \log m_0 \qquad p = q \tag{32}$$

We can now use that $m(X) = m(M) = m_0$, to obtain $m_{pq}(X)$ and $m_{qq}(X)$ as given in (7) and (10), respectively. Accordingly the functions g_{pq} and g are found to be

$$g_{pq}(\bullet) = (\bullet)^{\frac{1}{p-q}} \qquad p \neq q$$
$$g(\bullet) = \exp(\bullet) \qquad p = q \tag{33}$$

We have obtained axiomatizations of the generalized entropies and means which revealed, in particular, that the generalized entropies can be characterized as those entropies that cover the reference multiset case (the multiset of equal elements) and that are functions of only the ratio of the multisets' l^p and l^q norms. Similarly, the axiomatization also revealed that the generalized means can be characterized as those means which cover the reference multiset case and which are functions of

only the ratio of the multisets' p^{th} and q^{th} moments $\| X \|_p^p$ and $\| X \|_q^q$. We will now develop an axiomatization that links up with Section 2, yielding simultaneously a unique family of generalized entropies and means.

3.3. Unifying Axioms for Generalized Entropies and Means

We notice that, as is straightforward to verify:

$$\frac{\|X\|_p}{\|X\|_q} = \frac{\|Y\|_p}{\|Y\|_q} \text{ and } \frac{\|X\|_p^p}{\|X\|_q^q} = \frac{\|Y\|_p^p}{\|Y\|_q^q} \quad \Leftrightarrow \quad \|X\|_p = \|Y\|_p \text{ and } \|X\|_q = \|Y\|_q \tag{34}$$

This means that we can describe the generalized entropies and means also through a unifying set of axioms. To this end, let p, q be fixed real numbers obeying $pq \geq 0$. Consider maps S_{pq} and m_{pq}, which map multisets of positive real numbers into the real numbers. We call S_{pq} and m_{pq} generalized entropies and means of order p, q respectively, if they obey the following two axioms:

Unifying Axiom 1:

S_{pq} and m_{pq} applied to a multiset of k equal elements $M = \{m_0, m_0, \ldots, m_0\}$, $m_0 > 0$ yield the values $\log k$ and m_0, respectively.

Unifying axiom 2:

$$\|X\|_p = \|Y\|_p \text{ and } \|X\|_q = \|Y\|_q \quad \Rightarrow S_{pq}(X) := S_{pq}(Y) \text{ and } m_{pq}(X) := m_{pq}(Y) \quad p \neq q$$
$$\|X\|_q = \|Y\|_q \text{ and } \frac{\partial}{\partial q}\|X\|_q = \frac{\partial}{\partial q}\|Y\|_q \Rightarrow S_{qq}(X) := S_{qq}(Y) \text{ and } m_{qq}(X) := m_{qq}(Y) \quad p = q \tag{35}$$

Proposition: The maps S_{pq} and m_{pq} are unique and given by Equations (7)–(10).

Proof: The proofs are straightforward and proceed similarly to the proofs of the propositions related to the entropy and mean axioms.

4. Properties of S_{pq} and m_{pq}

We list in this section useful properties of m_{pq}, S_{pq} and k_{pq} with proofs in the appendix. The definitions of m_{pq}, S_{pq} and k_{pq} are given by (7)–(10). We also add two plots for an example multiset $X = \{10; 9; 8; 7; 6; 0.5; 0.4; 0.3; 0.2; 0.1\}$ in Figures 2 and 3 in order to provide some numerical illustration of the properties hereunder.

Figure 2. Numerical example showing the multiset elements $\{x_i\}$ *versus* their index i; with their corresponding mean m_{pq} and effective cardinality k_{pq} for different values of p, q.
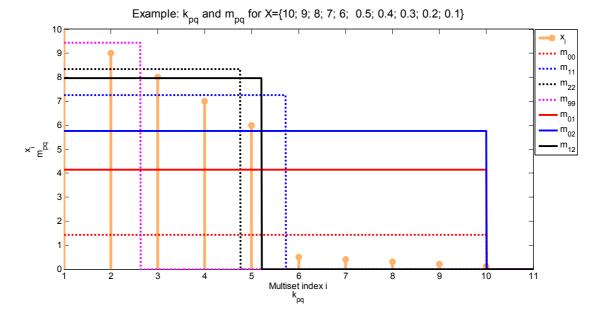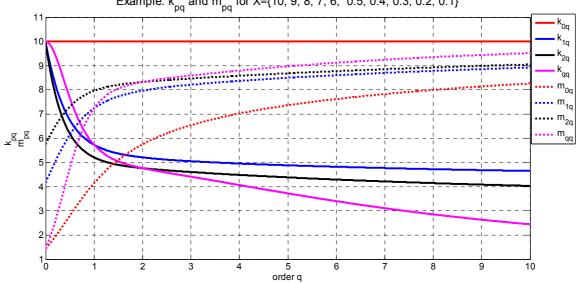

Figure 3. Numerical example for the multiset X showing its mean m_{pq} and effective cardinality k_{pq} plotted versus q for fixed values of p.

4.1. Scaling

Given a multiset $X = \{x_i : i = 1, 2, \ldots, K\}$ and a constant $\gamma > 0$, we have:

$$S_{pq}(\{\gamma x_i : i = 1, 2, \ldots, K\}) = S_{pq}(X)$$
$$m_{pq}(\{\gamma x_i : i = 1, 2, \ldots, K\}) = \gamma m_{pq}(X)$$

(36)

That is the entropy is invariant of the scaling, whereas the mean varies linearly with scaling. The proof is straightforward, based on (7).

4.2. Symmetry with Respect to the Elements of X

$S_{pq}(X)$ does not depend on the order of the elements of X.

4.3. Symmetry with Respect to the Order p,q

By exchanging p and q in (8), we readily find that:

$$S_{pq}(X) = S_{qp}(X) \tag{37}$$

4.4. Sign Change of the Order p,q

From (17), for $pq \geq 0$, we obtain:

$$S_{-p,-q}(\{x_i\}) = S_{p,q}(\{x_i^{-1} : x_i \neq 0\})$$
$$m_{-p,-q}(\{x_i\}) = m_{p,q}^{-1}(\{x_i^{-1} : x_i \neq 0\}) \tag{38}$$

4.5. Range of $S_{pq}(X)$

Let $N_X \geq 1$ be the number of the non-zero elements in X. Therefore:

$$0 \leq S_{pq}(X) \leq \log N_X \qquad pq \geq 0 \tag{39}$$

The minimum value occurs when X has exactly one non-zero element. For $p,q \neq 0$, the maximum value occurs when all the non-zero elements of X are equal. When either p or q is zero, $S_{pq}(X)$ yields the maximum value, $\log N_X$, for any distribution of X, which is physically intuitive since the zero[th] order renders all the non-zero multiset elements to an equal value and thus we reach the equiprobable case leading to maximum entropy. The proof of this property is in Appendix A.3.

4.6. Monotonicity of $S_{pq}(X)$ and $m_{pq}(X)$ with respect to p,q

We have the following results for the monotonicity of $S_{pq}(X)$ and $m_{pq}(X)$ with respect to p,q. The proofs are in Appendix B and Appendix C, respectively:

$$q\frac{\partial}{\partial p}S_{pq} \leq 0 \quad p \neq q, \ pq \geq 0$$
$$p\frac{\partial}{\partial p}S_{pp} \leq 0 \quad p \to q \tag{40}$$

$$\frac{\partial}{\partial p}m_{pq} \geq 0 \qquad p \neq q, \ pq \geq 0$$
$$\frac{\partial}{\partial p}m_{pp} \geq 0 \qquad p \to q \tag{41}$$

where equality holds when all the non-zero elements x_i are equal. Accordingly, for $p \neq q$ and $pq \geq 0$, by fixing one order (say q), m_{pq} is always non-decreasing with respect to the other order (p);

whereas S_{pq} is non-decreasing for $p < 0$, non-increasing for $p > 0$, with a maximum value at $p = 0$. The result is true when switching p, q from the symmetry Property 4.3.

Similarly, for the degenerate case $p \to q$, m_{pp} is always non-decreasing with respect to p; whereas S_{pp} is non-decreasing for $p < 0$, non-increasing for $p > 0$, with a maximum value at $p = 0$. In all cases, both the mean and entropy are invariant with respect to the order p, q if and only if all the non-zero elements x_i are equal.

This property explains the monotonicity of the curves in Figures 1 and 3.

4.7. Range of $m_{pq}(X)$

From Property 4.6, $m_{pq}(X)$ is non-decreasing with respect to p, q. From Table 1, we know that $m_{p,\infty}(X) = \max(X)$ and $m_{p,-\infty}(X) = \min(X)$. Therefore, $\min(X) \leq m_{pq}(X) \leq \max(X)$, which is an intuitive range for a mean. This is also true in the degenerate case $p \to q$, *i.e.*,

$$\min(X) \leq m_{pp}(X) \leq \max(X) \tag{42}$$

where, as usual, equality holds when all the non-zero elements x_i are equal.

4.8. Additivity of the Joint Multiset Entropy

For the two probability distribution multisets $X = \{x_i : i = 1, 2, \ldots, K_X\}$ and $Y = \{y_i : i = 1, 2, \ldots, K_Y\}$, we define the joint multiset $X * Y := \left\{ x_i y_j : \begin{array}{l} i = 1, 2, \ldots, K_X \\ j = 1, 2, \ldots, K_Y \end{array} \right\}$. Therefore, we have:

$$\begin{aligned} S(X * Y) &= S(X) + S(Y) \\ m(X * Y) &= m(X)m(Y) \end{aligned} \tag{43}$$

The proof is straightforward by using the fact that $\|X * Y\|_p = \|X\|_p \|Y\|_p$ along with (7)–(10), Property 4.8 is true for both $p \neq q$ and for the degenerate case $p \to q$.

4.9. Sub-Additivity of the Effective Cardinality Subject to the Multiset Additive Union Operation

Let $\uplus$ denote a multiset additive union operation [15] (page 50), e.g., $\{2,2\} \uplus \{1,1,2\} = \{1,1,2,2,2\}$. Let $X = \{x_i : i = 1, 2, \ldots, K_X\}$ and $Y = \{y_j : j = 1, 2, \ldots, K_Y\}$ be two multisets of non-negative real numbers. Moreover, let ξ and η be two positive real scaling factors of the elements of X and Y, respectively. Then:

$$k_{pq}(\{\xi x_i\} \uplus \{\eta y_j\}) \leq k_{pq}(X) + k_{pq}(Y) \tag{44}$$

where the equality holds under the following condition for the value ξ / η:

$$\frac{\xi_0}{\eta_0} = \begin{cases} \left(\dfrac{\sum_j y_j^q / \sum_j y_j^p}{\sum_i x_i^q / \sum_i x_i^p} \right)^{\frac{1}{q-p}} & p \neq q \\[4mm] \dfrac{\|Y\|_q}{\|X\|_q} \left(\dfrac{k_{qq}(X)}{k_{qq}(Y)} \right)^{1/q} & p \to q \end{cases} \tag{45}$$

This property is a generalization of the *effective alphabet size* of two disjoint alphabets mixture as discussed in [16] (Problem 2.10). To see this, set $p = q = 1$ and note that $\|X\|_1 = \|Y\|_1 = 1$ for a complete probability distribution. Accordingly, $k_{pq}(X)$ represents the *effective alphabet size* of X corresponding to the entropy $S_{pq}(X)$. The proof is in Appendix D.

4.10. Effective Rank of a Matrix

Let X be the multiset of the singular values of a matrix $\mathbf{M}$. Then, $k_{0q}(X) = \operatorname{rank}(\mathbf{M})$. Accordingly, for general p, q, $k_{pq}(X)$ can be viewed as a biased effective rank of $\mathbf{M}$ corresponding to the order p, q. From Property 4.5, we have $1 \le k_{pq}(X) \le \operatorname{rank}(\mathbf{M})$. The minimum value occurs when $\mathbf{M}$ has exactly one non-zero singular value (the rank of $\mathbf{M}$ is 1). The maximum value, $\operatorname{rank}(\mathbf{M})$, is reached for any $pq > 0$ when all the non-zero singular values are equal. The effective rank can be helpful to determine, in a well-defined p, q sense, how to view an ill-conditioned matrix, which is a full-rank from a mathematical perspective, but is effectively behaving as if possessing a lower rank. Such ill-conditioned matrices often arise in problems of oversampling or determining the degrees of freedom, where the singular values, ordered in non-increasing order by definition, exhibit some sort of "knee cut-off", similar to $\{x_i\}$ in Figure 2. A biased effective rank can help to compare matrices when the knee cut-off is not sharp, thus giving more weight to the small or large singular values according to the order p, q in a consistent manner for different matrices. An example thereof is the evaluation of the degrees of freedom of some applications such as multi-antenna systems [3–6], optical imaging systems [20], or in general any case of similar limitation to space-bandwidth product [21].

4.11. Geometrical Interpretation of S_{pq} and m_{pq} on Log Scale, with Thermodynamics Analogy

For the multiset $X = \{x_i\}$, after taking the logarithm of both sides of (5), we obtain a simple relation between $\|X\|_q$, S_{pq} and m_{pq} as in (11):

$$\begin{aligned}
\log \|X\|_q^q &= \log \|M\|_q^q = S_{pq} + q \log m_{pq} \\
\log \|X\|_p^p &= \log \|M\|_p^p = S_{pq} + p \log m_{pq}
\end{aligned} \tag{46}$$

Accordingly, a secant cutting the function $f(q) = \log \|X\|_q^q$ at $q = p_0$ and $q = q_0$ will have a slope and intercept of $\log m_{p_0 q_0}$ and $S_{p_0 q_0}$, respectively, as shown in Figure 4. Based on the discussion of Section 1, (46) readily yields the following analogous expressions for a thermodynamics system described by the Boltzmann factors $\{e^{-E_i}\}$:

$$\begin{aligned}
\log Z_{\beta 1}(X) &= S_{\beta 1, \beta 2} - \beta_1 E_0 \\
\log Z_{\beta 2}(X) &= S_{\beta 1, \beta 2} - \beta_2 E_0
\end{aligned} \tag{47}$$

In the limiting case, when $p \to q, \beta_1 \to \beta_2$, the secant in Figure 4 becomes a tangent and we get the Gibbs–Shannon entropy at the inverse temperature β. Consequently, (47) can be re-written after introducing the Boltzmann constant K_B and using the absolute temperature $T = 1 / K_B \beta$ as:

$$-K_B T \log Z_\beta := A_H$$
$$= E_0 - TS_\beta$$

(48)

where A_H is the Helmholtz free energy of the system.

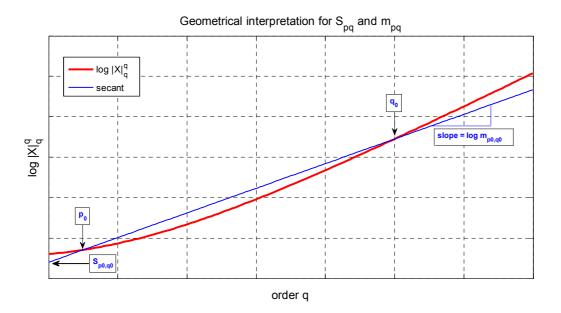**Figure 4.** A secant of $\log \|X\|_q^q$ *versus* q.

Geometrical interpretation for S_{pq} and m_{pq}

5. Discussion and Conclusions

A two-parameter family of cardinalities, entropies and means has been derived for multisets of non-negative elements. Rather than starting from thermodynamic or information theoretic considerations to derive entropies and means, see e.g., [9,17–19], we here defined the generalized entropies and means through simple abstract axioms. There are other families of entropies in the literature (e.g., [23]), which are a generalization of Shannon entropy. The generalized entropy in this manuscript is shown to preserve the additivity (Property 4.8), which is not the case with the generalized entropies based on Tsallis non-additive entropy as in [23].

Our first two axiomatizations treat the generalized entropies and means separately. It revealed that the generalized entropies are exactly those entropies that are functions of only the ratio of the multisets' l^p and l^q *norms*. It also revealed that the generalized means are exactly those means that are functions only of the ratio of the multisets' p^{th} and q^{th} *moments*, $\| X \|_p^p$ and $\| X \|_q^q$. Subsequently, our unifying axiomatization characterized the generalized entropies and means together. This showed that if two multisets have exactly the same l^p and l^q norms, then they share the same generalized entropy and mean.

We presented several key features of the new families of generalized entropies and means, for example, that the family of generalized entropies contains and generalizes the Rényi family of entropies, of which the Shannon entropy is a special case, thus including some of the desiderata for entropies [22]. We also showed the monotonicity with respect to p, q, extreme values, symmetry with respect to p, q, and additivity preservation. The effective cardinality k_{pq} measures the distribution

uniformity of the multiset elements in the sense of the *p*- and *q*-norm equivalence to a reference flat multiset. From an information theory perspective, S_{pq} and k_{pq} represent a two-parameter entropy of order p, q and its corresponding effective alphabet size, respectively, when a probability distribution is constructed after proper normalization of the multiset elements. Furthermore, we recall that knowing the ℓ^p and ℓ^q norms of a multiset is to know the multiset's p^{th} and q^{th} moments. Our findings here therefore imply that knowledge of a multiset's p^{th} and q^{th} moments is exactly enough information to deduce the multiset's (*p*,*q*) entropy and (*p*,*q*) mean. Further, knowledge of sufficiently many moments of a multiset can be sufficient to reconstruct the multiset. Conversely, it should be interesting to examine how many (*p*,*q*)-entropies and/or (*p*,*q*)-means are required to completely determine the multiset.

Regarding the thermodynamic interpretation, we noticed that to require that the ℓ^p and ℓ^q norms of multisets coincide is mathematically equivalent to requiring that the two partition functions of two thermodynamic systems coincide at two temperatures. This in turn is equivalent to requiring that the Helmholtz free energy of the two thermodynamic systems coincide at two temperatures. The Helmholtz free energy represents the maximum mechanical work that can be extracted from a thermodynamic system under certain idealized circumstances. This suggests that there perhaps exists a thermodynamic interpretation of the generalized entropies and means in terms of the extractability of mechanical work. In this case, the fact that the generalized entropies and means depend on two rather than one temperature could be related to the fact that the maximum efficiency of a heat engine, obtained in Carnot cycles, is a function of two temperatures. We did show that in the limiting case, when the two temperatures become the same, one recovers the usual Boltzmann factor weighted *mean energy* as well as the usual Shannon/von Neumann *entropy*.

Acknowledgments

References

1. Rényi, A. On the foundations of information theory. *Rev. Int. Stat. Inst.* **1965**, *33*, 1–14.
2. Beck C. Generalized information and entropy measures in physics. *Contemp. Phys.* **2009**, *50*, 495–510.
3. Poon, A.S.Y.; Brodersen, R.W.; Tse, D.N.C. Degrees of freedom in multiple-antenna channels: A signal space approach. *IEEE Trans. Inform. Theor.* **2005**, *51*, 523–536.
4. Migliore, M.D. On the role of the number of degrees of freedom of the field in MIMO channels. *IEEE Trans. Antenn. Propag.* **2006**, *54*, 620–628.
5. Elnaggar, M.S. Electromagnetic Dimensionality of Deterministic Multi-Polarization MIMO Systems. Ph.D. Thesis, University of Waterloo, Waterloo, ON, Canada, 2007; Available online: http://uwspace.uwaterloo.ca/handle/10012/3434 (accessed on 18 April 2012).

6.   Elnaggar, M.S.; Safavi-Naeini, S.; Chaudhuri, S.K. A novel dimensionality metric for multi-antenna systems. In Proceedings of Asia-Pacific Microwave Conference (APMC 2006), Yokohama, Japan, December 2006; pp. 242–245.

7.   Elnaggar, M.S.; Kempf, A. On a Generic Entropy Measure in Physics and Information. In Proceedings of the 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT 2009), Seoul, Korea, June 2009.

8.   Beck, C.; Schlogl, F. *Thermodynamics of Chaotic Systems: An Introduction*; Cambridge University Press: Cambridge, UK, 1995.

9.   Rényi, A. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability, University of California, Berkeley, CA, USA, 1961; Volume 1, pp. 541–561.

10. Aczél, J.; Forte, B.; Ng, C.T. Why the Shannon and Hartley entropies are 'natural'. *Adv. Appl. Probab.* **1974**, *6*, 131–146.

11. Hartley, R.V.L. Transmission of information. *Bell Syst. Tech. J.* **1928**, *7*, 535–564.

12. Rényi, A. *Probability Theory*; North-Holland Pub. Co.: Amsterdam, The Netherlands, 1970.

13. Klir, G.J.; Folger, T.A. *Fuzzy Sets, Uncertainty and Information*; Prentice Hall: Upper Saddle River, NJ, USA, 1988.

14. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Dover: New York, NY, USA, 1957.

15. Blizard, W.D. Multiset theory. *Notre Dame J. Formal Logic* **1989**, *30*, 36–66.

16. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 1991.

17. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.

18. Ash, R.B. *Information Theory*; Interscience Publishers: New York, NY, USA, 1965.

19. Reza, F.M. *An Introduction to Information Theory*. Dover: New York, NY, USA, 1994.

20. Gori, F.; Guattari, G. Shannon number and degrees of freedom of an image. *Opt. Commun.* **1973**, *7*, 163–165.

21. Landau, H.; Pollak, H. Prolate spheroidal wave functions, Fourier analysis and uncertainty: Part III: The dimension of the space of essentially time- and band-limited signals. *Bell Syst. Tech. J.* **1962**, *41*, 1295–1336.

22. Aczel, J.; Daroczy, Z. On Measures of Information and Their Characterizations; Academic Press: New York, NY, USA, 1975.

23. Shigeru, F. An axiomatic characterization of a two-parameter extended relative entropy. *J. Math. Phys.* **2010**, *51*, 123302.

Appendix A. Proof of Property 4.5: Range of S_{pq}

A.1. Non-Increasing $\|X\|_p$ with Respect to p

We want to show that $\|X\|_p$ is non-increasing with respect to p. This is equivalent to showing that

$$\frac{\partial}{\partial p}(\log\|X\|_p) = \frac{\partial}{\partial p}\left(\frac{\log\sum_{i=1}^{N}x_i^p}{p}\right) \le 0.$$

$$p^2\frac{\partial}{\partial p}(\log\|X\|_p) = p\sum_i\frac{x_i^p}{\sum_j x_j^p}\log x_i - \log\sum_j x_j^p$$

$$= \sum_i\frac{x_i^p}{\sum_j x_j^p}\log x_i^p - \sum_i\frac{x_i^p}{\sum_j x_j^p}\log\sum_j x_j^p$$

$$= \sum_i\frac{x_i^p}{\sum_j x_j^p}\log\frac{x_i^p}{\sum_j x_j^p} \le 0$$

Therefore, $\|X\|_p \ge \|X\|_q$ for $p<q$. The equality occurs only when $x_i^p = \sum x_i^p$, *i.e.*, when $\{x_i\}$ includes exactly one non-zero element. Note that there is a singularity for $\log\|X\|_p$ such that $\lim_{p\to 0^+}\log\|X\|_p = \infty$ and $\lim_{p\to 0^-}\log\|X\|_p = -\infty$. The property herein applies over the intervals $-\infty < p < 0$ and $0 < p < \infty$.

A.2. Non-Decreasing Generalized p-Mean with Respect to p

Let X be a multiset of N strictly positive numbers. For $p < q$, we want to show that $\left(\frac{\sum_{i=1}^{N}x_i^p}{N}\right)^{1/p} \le \left(\frac{\sum_{i=1}^{N}x_i^q}{N}\right)^{1/q}$. To this end, we define $f(x)=x^{q/p}$, where $x>0$ and $p/q>0$, *i.e.*, $p,q\ne 0$ have the same sign. Clearly, $\frac{d^2}{dx^2}f(x)>0$ and thus $f(x)$ is convex. From Jensen's inequality,

$$f\left(\sum w_i x_i\right) \le \sum w_i f(x_i), \text{ where } \sum_{i=1}^{n}w_i = 1. \text{ Accordingly, } f\left(\sum w_i x_i^p\right) \le \sum w_i f(x_i^p), \text{ hence}$$

$$\left(\sum w_i x_i^p\right)^{q/p} \le \sum w_i(x_i^p)^{q/p} = \sum w_i x_i^q$$

$$\left(\sum w_i x_i^p\right)^{1/p} \le \left(\sum w_i x_i^q\right)^{1/q}$$

By setting $w_i = 1/N$, the proof is complete. Equality holds when all x_i are equal.

A.3. Proof of Property 4.5: $0 \le S_{pq} \le \log N$

We start by proving that $S_{pq} \ge 0$. From Lemma A.1, for $p < q$, we have $\log\|X\|_p \ge \log\|X\|_q$. Therefore, $\log\|X\|_p - \log\|X\|_q \ge 0$, $p, q \ne 0$ and have the same sign. Since $1/p - 1/q > 0$, we get $\dfrac{\log\|X\|_p - \log\|X\|_q}{1/p - 1/q} = S_{pq} \ge 0$. Following the same steps for $p > q$ yields the same result. Finally, when $p \to q$, we have $S_{qq} = -\sum_i \dfrac{x_i^q}{\sum_j x_j^q} \log\left(\dfrac{x_i^q}{\sum_j x_j^q}\right) \ge 0$. In all cases, the equality holds when $\{x_i\}$ includes exactly one non-zero element.

Next, we show that $S_{pq} \le \log N$. From Lemma A.2, assuming both $p, q \ne 0$ have the same sign, for $p < q$, we have $\left(\dfrac{\sum_{i=1}^N x_i^p}{N}\right)^{1/p} \le \left(\dfrac{\sum_{i=1}^N x_i^q}{N}\right)^{1/q}$. Taking the logarithm yields $\dfrac{\log\|X\|_p - \log\|X\|_q}{1/p - 1/q} = S_{pq} \le \log N$. For $p > q$, we obtain the same result.

When $p \to q$, we have:

$$S_{qq} - \log N = -\sum_{i=1}^N \frac{x_i^q}{\sum_{j=1}^N x_j^q} \log\left(\frac{x_i^q}{\sum_{j=1}^N x_j^q}\right) - \sum_{i=1}^N \frac{x_i^q}{\sum_{j=1}^N x_j^q} \log N$$

$$= \sum_{i=1}^N \frac{x_i^q}{\sum_{j=1}^N x_j^q} \log\left(\frac{\sum_{j=1}^N x_j^q}{N x_i^q}\right)$$

$$\le \sum_{i=1}^N \frac{x_i^q}{\sum_{j=1}^N x_j^q}\left(\frac{\sum_{j=1}^N x_j^q}{N x_i^q} - 1\right) = \sum_{i=1}^N \left(\frac{1}{N} - \frac{x_i^q}{\sum_{j=1}^N x_j^q}\right) = 0$$

where $\log x \le (x-1)$ was used. In all cases, equality holds when all the non-zero elements of X are equal.

Appendix B. Proof of Property 4.6: Monotonicity of S_{pq}

B.1. Non-Degenerate Case $p \neq q$

We have $S_{pq} = \log k_{pq} = \dfrac{q \log \sum_i x_i^p - p \log \sum_i x_i^q}{q - p}$. Accordingly:

$$(q-p)^2 \frac{\partial}{\partial p} S_{pq} = q(q-p)\sum_i \frac{x_i^p}{\sum_j x_j^p} \log x_i - (q-p)\log \sum_i x_i^q + q\log \sum_i x_i^p - p\log \sum_i x_i^q$$

$$= q(q-p)\sum_i \frac{x_i^p}{\sum_j x_j^p} \log x_i - q\log \sum_i x_i^q + q\log \sum_i x_i^p$$

$$= q(q-p)\sum_i \frac{x_i^p}{\sum_j x_j^p} \log x_i - q\log \sum_j x_j^q + q\log \sum_j x_j^p$$

where we inserted another summation index for the last two terms in the last step. Therefore:

$$\frac{(q-p)^2}{q} \frac{\partial}{\partial p} S_{pq} = (q-p)\sum_i \frac{x_i^p}{\sum_j x_j^p} \log x_i - \log \sum_j x_j^q + \log \sum_j x_j^p$$

$$= \sum_i \frac{x_i^p}{\sum_j x_j^p} \log x_i^q - \sum_i \frac{x_i^p}{\sum_j x_j^p} \log x_i^p - \log \sum_j x_j^q + \log \sum_j x_j^p$$

$$= \sum_i \frac{x_i^p}{\sum_j x_j^p} \log\left(\frac{x_i^q \sum_j x_j^p}{x_i^p \sum_j x_j^q} \right)$$

where we have multiplied the last two terms in the second step by $\dfrac{\sum_i x_i^p}{\sum_j x_j^p}$. Consequently:

$$\frac{(q-p)^2}{q} \frac{\partial}{\partial p} S_{pq} \leq \sum_i \frac{x_i^p}{\sum_j x_j^p} \left(\frac{x_i^q \sum_j x_j^p}{x_i^p \sum_j x_j^q} - 1 \right) = 0 ,$$

where we have used the fact that $\log x \leq (x-1)$. The equality occurs only when $\dfrac{x_i^q}{\sum_j x_j^q} = \dfrac{x_i^p}{\sum_j x_j^p}$ for

every x_i, implying that each x_i is either zero or some fixed value x_0.

Accordingly, after multiplying the last inequality by $\dfrac{q^2}{(q-p)^2}$, we obtain:

$$q \frac{\partial}{\partial p} S_{pq} \leq 0 \quad p \neq q ,$$

where the equality holds when all the non-zero elements x_i are equal.

Since p and q are assumed to have the same sign (from the condition $pq \geq 0$), we deduce that, with respect to either p or q (while fixing the other order), S_{pq} is non-decreasing for negative order p,q,

non-increasing for positive order p,q, with a maximum value at either $p,q = 0$ (note that $\left.\frac{\partial}{\partial p} S_{pq}\right|_{q=0} = 0$). S_{pq} is invariant to p,q if and only if all the non-zero elements x_i are equal.

B.2. Degenerate Case $p \to q$

From Appendix B.1, if $0 \le a \le b \le c \le d$, then $S_{cd} \le S_{bd} \le S_{bc} \le S_{ac} \le S_{ab}$. In the limiting case, when $c \to d^-$ and $b \to a^+$, and knowing from (9) that this limit exists, we readily obtain $S_{dd} \le S_{aa}$. In a similar fashion, for negative indices values $d \le c \le b \le a \le 0$, we get $S_{aa} \ge S_{dd}$. In any case, the equality occurs when all the non-zero elements x_i are equal. S_{pp} has a maximum value when $p = 0$. Accordingly:

$$p \frac{\partial}{\partial p} S_{pp} \le 0 .$$

Appendix C. Proof of Property 4.6: Monotonicity of m_{pq}

C.1. Non-Degenerate Case $p \ne q$

We have from (5) $\log \|X\|_q = \log m_{pq} + S_{pq} / q$. Differentiating with respect to p, we get:

$$\frac{\partial}{\partial p} m_{pq} = -m_{pq} \frac{\frac{\partial}{\partial p} S_{pq}}{q} .$$

Since $k_{pq} > 0$ (Property 4.5), therefore (5) confirms that $m_{pq} > 0$ as well because the norm is defined to discard any zero elements (6). From Appendix B, $\frac{\frac{\partial}{\partial p} S_{pq}}{q} \le 0$. Therefore:

$$\frac{\partial}{\partial p} m_{pq} \ge 0 \quad p,q \ne 0 \quad p \ne q ,$$

with equality when all the non-zero elements x_i are equal.

C.2. Degenerate Case $p \to q$

From (9), $\log m_{pp}(X) = \log \|X\|_p - S_{pp}(X) / p$, accordingly:

$$\frac{\frac{\partial}{\partial p} m_{pp}}{m_{pp}} = \frac{\frac{\partial}{\partial p} \|X\|_p}{\|X\|_p} - \frac{p \frac{\partial}{\partial p} S_{pp} - S_{pp}}{p^2}$$

Moreover, from (9), $S_{qq}(X) = -q^2 \dfrac{\dfrac{\partial}{\partial q}\|X\|_q}{\|X\|_q}$. Consequently, $p^2 \dfrac{\dfrac{\partial}{\partial p} m_{pp}}{m_{pp}} = -p \dfrac{\partial}{\partial p} S_{pp} \geq 0$, where we used the result of Appendix B.2. Therefore, since $m_{pp} > 0$, we get:

$$\frac{\partial}{\partial p} m_{pp} \geq 0.$$

Appendix D. Proof of Property 4.9: Sub-Additivity of the Effective Cardinality

For $p \neq q$, we want to show that $k_{pq}(\{\xi x_i\} \uplus \{\eta y_j\}) \leq k_{pq}(X) + k_{pq}(Y)$, where $\xi, \eta \neq 0$. Our objective is to find a condition in terms of ξ and η that maximizes $S_{pq}(\{\xi x_i\} \uplus \{\eta y_j\})$. We have:

$$(p-q)S_{pq}(\{\xi x_i\} \uplus \{\eta y_j\}) = p \log\left(\sum_i \xi^q x_i^q + \sum_j \eta^q y_j^q\right) - q \log\left(\sum_i \xi^p x_i^p + \sum_j \eta^p y_j^p\right) \tag{49}$$

By setting $\dfrac{\partial}{\partial \xi} S_{pq} = 0$ and $\dfrac{\partial}{\partial \eta} S_{pq} = 0$, we obtain the following two conditions:

$$\frac{\sum_i \xi_0^q x_i^q}{\sum_i \xi_0^q x_i^q + \sum_j \eta_0^q y_j^q} = \frac{\sum_i \xi_0^p x_i^p}{\sum_i \xi_0^p x_i^p + \sum_j \eta_0^p y_j^p} \tag{50}$$

$$\frac{\sum_j \eta_0^q y_j^q}{\sum_i \xi_0^q x_i^q + \sum_j \eta_0^q y_j^q} = \frac{\sum_j \eta_0^p y_j^p}{\sum_i \xi_0^p x_i^p + \sum_j \eta_0^p y_j^p} \tag{51}$$

From (50) and (51), we have:

$$\frac{\sum_i \xi_0^q x_i^q}{\sum_i \xi_0^p x_i^p} = \frac{\sum_j \eta_0^q y_j^q}{\sum_j \eta_0^p y_j^p} \tag{52}$$

Accordingly, the required proportionality condition is:

$$\frac{\xi_0}{\eta_0} = \left(\frac{\sum_j y_j^q / \sum_j y_j^p}{\sum_i x_i^q / \sum_i x_i^p}\right)^{\frac{1}{q-p}} \tag{53}$$

From (49), we have:

$$k_{pq}(\{\xi_0 x_i\} \uplus \{\eta_0 y_j\}) = \left(\frac{\left(\sum_i \xi_0^q x_i^q + \sum_j \eta_0^q y_j^q\right)^p}{\left(\sum_i \xi_0^p x_i^p + \sum_j \eta_0^p y_j^p\right)^q}\right)^{\frac{1}{p-q}} \tag{54}$$

Substituting $\sum_i \xi_0{}^p x_i{}^p + \sum_j \eta_0{}^p y_j{}^p$ from (50) in (54) yields:

$$k_{pq}(\{\xi_0 x_i\} \uplus \{\eta_0 y_j\}) = \left(\frac{\left(\sum_i \xi_0{}^q x_i{}^q\right)^q}{\left(\sum_i \xi_0{}^p x_i{}^p\right)^q} \left(\sum_i \xi_0{}^q x_i{}^q + \sum_j \eta_0{}^q y_j{}^q\right)^{p-q} \right)^{\frac{1}{p-q}}$$

$$= \frac{\left(\sum_i \xi_0{}^q x_i{}^q\right)^{\frac{q}{p-q}}}{\left(\sum_i \xi_0{}^p x_i{}^p\right)^{\frac{q}{p-q}}} \left(\sum_i \xi_0{}^q x_i{}^q + \sum_j \eta_0{}^q y_j{}^q\right)$$

$$= \frac{\left(\sum_i \xi_0{}^q x_i{}^q\right)^{1+\frac{q}{p-q}}}{\left(\sum_i \xi_0{}^p x_i{}^p\right)^{\frac{q}{p-q}}} + \left(\sum_j \eta_0{}^q y_j{}^q\right)\left(\frac{\sum_i \xi_0{}^q x_i{}^q}{\sum_i \xi_0{}^p x_i{}^p}\right)^{\frac{q}{p-q}}$$

$$= \frac{\left(\sum_i \xi_0{}^q x_i{}^q\right)^{\frac{p}{p-q}}}{\left(\sum_i \xi_0{}^p x_i{}^p\right)^{\frac{q}{p-q}}} + \left(\sum_j \eta_0{}^q y_j{}^q\right)\left(\frac{\sum_j \eta_0{}^q y_j{}^q}{\sum_j \eta_0{}^p y_j{}^p}\right)^{\frac{q}{p-q}}$$

$$= \frac{\left(\sum_i \xi_0{}^q x_i{}^q\right)^{\frac{p}{p-q}}}{\left(\sum_i \xi_0{}^p x_i{}^p\right)^{\frac{q}{p-q}}} + \frac{\left(\sum_j \eta_0{}^q y_j{}^q\right)^{\frac{p}{p-q}}}{\left(\sum_j \eta_0{}^p y_j{}^p\right)^{\frac{q}{p-q}}}$$

where we employed (52) to obtain the second to last step. Consequently, since ξ_0 and η_0 cancel out in the last step, we get:

$$k_{pq}(\xi_0 X \uplus \eta_0 Y) = \left(\frac{\left(\sum_i x_i{}^q\right)^p}{\left(\sum_i x_i{}^p\right)^q} \right)^{\frac{1}{p-q}} + \left(\frac{\left(\sum_j y_j{}^q\right)^p}{\left(\sum_j y_j{}^p\right)^q} \right)^{\frac{1}{p-q}} .$$

$$= k_{pq}(X) + k_{pq}(Y)$$

In order to confirm that (53) is indeed a maximizing condition of $k_{pq}(\xi_0 X \uplus \eta_0 Y)$, consider the following example. Let $X = \{x_0\}$ and $Y = \{y_0\}$ with multiplicity K_X and K_Y, respectively. From the range of k_{pq} (Property 4.5), we know that $\max(k_{pq}(\xi_0 X \uplus \eta_0 Y)) = K_X + K_Y$ only when all the elements are equal, *i.e.*, when $\xi_0 x_0 = \eta_0 y_0$. Indeed, (53) confirms this maximizing condition and thus $k_{pq}(\xi X \uplus \eta Y) \le k_{pq}(X) + k_{pq}(Y)$.

By taking the limit when $p \to q$, (53) becomes

$$\frac{\xi_0}{\eta_0} = \frac{\|Y\|_q}{\|X\|_q} \left(\frac{k_{qq}(X)}{k_{qq}(Y)} \right)^{1/q} \qquad p \to q \qquad (55)$$

and $k_{qq}(\xi X \uplus \eta Y) \le k_{qq}(X) + k_{qq}(Y)$.