Entropy **2011**, *13*, 1945-1957; doi:10.3390/e13111945

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

A Characterization of Entropy in Terms of Information Loss

John C. Baez [1,2], Tobias Fritz [3,⋆] and Tom Leinster [4]

[1] Department of Mathematics, University of California, Riverside, CA 92521, USA;
E-Mail: baez@math.ucr.edu

[2] Centre for Quantum Technologies, National University of Singapore, 117543, Singapore

[3] Institut de Ciències Fotòniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain

[4] School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QW, UK;
E-Mail: tom.leinster@glasgow.ac.uk

⋆ Author to whom correspondence should be addressed; E-Mail: tobias.fritz@icfo.es.

Received: 11 October 2011; in revised form: 18 November 2011 / Accepted: 21 November 2011 / Published: 24 November 2011

Abstract: There are numerous characterizations of Shannon entropy and Tsallis entropy as measures of information obeying certain properties. Using work by Faddeev and Furuichi, we derive a very simple characterization. Instead of focusing on the entropy of a probability measure on a finite set, this characterization focuses on the "information loss", or change in entropy, associated with a measure-preserving function. Information loss is a special case of conditional entropy: namely, it is the entropy of a random variable conditioned on some function of that variable. We show that Shannon entropy gives the only concept of information loss that is functorial, convex-linear and continuous. This characterization naturally generalizes to Tsallis entropy as well.

Keywords: Shannon entropy; Tsallis entropy; information theory; measure-preserving function

Classification: MSC Primary: 94A17, Secondary: 62B10

---

1. Introduction

The Shannon entropy [1] of a probability measure p on a finite set X is given by:

$$H(p) = -\sum_{i \in X} p_i \ln(p_i)$$

There are many theorems that seek to characterize Shannon entropy starting from plausible assumptions; see for example the book by Aczél and Daróczy [2]. Here we give a new and very simple characterization theorem. The main novelty is that we do not focus directly on the entropy of a single probability measure, but rather, on the *change* in entropy associated with a measure-preserving function. The entropy of a single probability measure can be recovered as the change in entropy of the unique measure-preserving function onto the one-point space.

A measure-preserving function can map several points to the same point, but not vice versa, so this change in entropy is always a *decrease*. Since the second law of thermodynamics speaks of entropy *increase*, this may seem counterintuitive. It may seem less so if we think of the function as some kind of data processing that does not introduce any additional randomness. Then the entropy can only decrease, and we can talk about the "information loss" associated with the function.

Some examples may help to clarify this point. Consider the only possible map $f\colon \{a, b\} \to \{c\}$. Suppose p is the probability measure on $\{a, b\}$ such that each point has measure $1/2$, while q is the unique probability measure on the set $\{c\}$. Then $H(p) = \ln 2$, while $H(q) = 0$. The information loss associated with the map f is defined to be $H(p) - H(q)$, which in this case equals $\ln 2$. In other words, the measure-preserving map f loses one bit of information.

On the other hand, f is also measure-preserving if we replace p by the probability measure p' for which a has measure 1 and b has measure 0. Since $H(p') = 0$, the function f now has information loss $H(p') - H(q) = 0$. It may seem odd to say that f loses no information: after all, it maps a and b to the the same point. However, because the point b has probability zero with respect to p', knowing that $f(x) = c$ lets us conclude that $x = a$ with probability one.

The shift in emphasis from probability measures to measure-preserving *functions* suggests that it will be useful to adopt the perspective of category theory [3], where one has objects and *morphisms* between them. However, the reader need only know the definition of "category" to understand this paper.

Our main result is that Shannon entropy has a very simple characterization in terms of information loss. To state it, we consider a category where a morphism $f\colon p \to q$ is a measure-preserving function between finite sets equipped with probability measures. We assume F is a function that assigns to any such morphism a number $F(f) \in [0, \infty)$, which we call its **information loss**. We also assume that F obeys three axioms. If we call a morphism a "process" (to be thought of as deterministic), we can state these roughly in words as follows. For the precise statement, including all the definitions, see Section 2.

(i) **Functoriality**. Given a process consisting of two stages, the amount of information lost in the whole process is the sum of the amounts lost at each stage:

$$F(f \circ g) = F(f) + F(g)$$

(ii) **Convex linearity**. If we flip a probability-λ coin to decide whether to do one process or another, the information lost is λ times the information lost by the first process plus $(1-\lambda)$ times the information lost by the second:

$$F(\lambda f \oplus (1-\lambda)g) = \lambda F(f) + (1-\lambda)F(g)$$

(iii) **Continuity**. If we change a process slightly, the information lost changes only slightly: $F(f)$ is a continuous function of f.

Given these assumptions, we conclude that there exists a constant $c \geq 0$ such that for any $f\colon p \to q$, we have

$$F(f) = c(H(p) - H(q))$$

The charm of this result is that the first two hypotheses look like linear conditions, and none of the hypotheses hint at any special role for the function $-p \ln p$, but it emerges in the conclusion. The key here is a result of Faddeev [4] described in Section 4.

For many scientific purposes, *probability* measures are not enough. Our result extends to general measures on finite sets, as follows. Any measure on a finite set can be expressed as λp for some scalar λ and probability measure p, and we define $H(\lambda p) = \lambda H(p)$. In this more general setting, we are no longer confined to taking *convex* linear combinations of measures. Accordingly, the convex linearity condition in our main theorem is replaced by two conditions: additivity ($F(f \oplus g) = F(f) + F(g)$) and homogeneity ($F(\lambda f) = \lambda F(f)$). As before, the conclusion is that, up to a multiplicative constant, F assigns to each morphism $f\colon p \to q$ the information loss $H(p) - H(q)$.

It is natural to wonder what happens when we replace the homogeneity axiom $F(\lambda f) = \lambda F(f)$ by a more general homogeneity condition:

$$F(\lambda f) = \lambda^\alpha F(f)$$

for some number $\alpha > 0$. In this case we find that $F(f)$ is proportional to $H_\alpha(p) - H_\alpha(q)$, where H_α is the so-called Tsallis entropy of order α.

2. The Main Result

We work with finite sets equipped with probability measures. All measures on a finite set X will be assumed nonnegative and defined on the σ-algebra of all subsets of X. Any such measure is determined by its values on singletons, so we will think of a probability measure p on X as an X-tuple of numbers $p_i \in [0,1]$ ($i \in X$) satisfying $\sum p_i = 1$.

Definition 1. *Let* `FinProb` *be the category where an object (X,p) is given by a finite set X equipped with a probability measure p, and where a morphism $f\colon (X,p) \to (Y,q)$ is a measure-preserving function from (X,p) to (Y,q), that is, a function $f\colon X \to Y$ such that*

$$q_j = \sum_{i \in f^{-1}(j)} p_i$$

for all $j \in Y$.

We will usually write an object (X, p) as p for short, and write a morphism $f\colon (X, p) \to (Y, q)$ as simply $f : p \to q$.

There is a way to take convex linear combinations of objects and morphisms in `FinProb`. Let (X, p) and (Y, q) be finite sets equipped with probability measures, and let $\lambda \in [0, 1]$. Then there is a probability measure

$$\lambda p \oplus (1 - \lambda)q$$

on the disjoint union of the sets X and Y, whose value at a point k is given by

$$(\lambda p \oplus (1 - \lambda)q)_k = \begin{cases} \lambda p_k & \text{if } k \in X \\ (1 - \lambda)q_k & \text{if } k \in Y \end{cases}$$

Given morphisms $f : p \to p'$ and $g : q \to q'$, there is a unique morphism

$$\lambda f \oplus (1 - \lambda)g : \lambda p \oplus (1 - \lambda)q \longrightarrow \lambda p' \oplus (1 - \lambda)q'$$

that restricts to f on the measure space p and to g on the measure space q.

The same notation can be extended, in the obvious way, to convex combinations of more than two objects or morphisms. For example, given objects $p(1), \ldots, p(n)$ of `FinProb` and nonnegative scalars $\lambda_1, \ldots, \lambda_n$ summing to 1, there is a new object $\bigoplus_{i=1}^n \lambda_i p(i)$.

Recall that the **Shannon entropy** of a probability measure p on a finite set X is

$$H(p) = -\sum_{i \in X} p_i \ln(p_i) \in [0, \infty)$$

with the convention that $0 \ln(0) = 0$.

Theorem 2. *Suppose F is any map sending morphisms in* `FinProb` *to numbers in $[0, \infty)$ and obeying these three axioms:*

(i) ***Functoriality****:*

$$F(f \circ g) = F(f) + F(g) \tag{1}$$

whenever f, g are composable morphisms.

(ii) ***Convex linearity****:*

$$F(\lambda f \oplus (1 - \lambda)g) = \lambda F(f) + (1 - \lambda)F(g) \tag{2}$$

for all morphisms f, g and scalars $\lambda \in [0, 1]$.

(iii) ***Continuity****: F is continuous.*

Then there exists a constant $c \geq 0$ such that for any morphism $f : p \to q$ in `FinProb`*,*

$$F(f) = c(H(p) - H(q))$$

where $H(p)$ is the Shannon entropy of p. Conversely, for any constant $c \geq 0$, this formula determines a map F obeying Conditions (i)–(iii).

We need to explain Condition (iii). A sequence of morphisms

$$(X_n, p(n)) \xrightarrow{f_n} (Y_n, q(n))$$

in `FinProb` **converges** to a morphism $(X, p) \xrightarrow{f} (Y, q)$ if:

- for all sufficiently large n, we have $X_n = X$, $Y_n = Y$, and $f_n(i) = f(i)$ for all $i \in X$;
- $p(n) \to p$ and $q(n) \to q$ pointwise.

We define F to be **continuous** if $F(f_n) \to F(f)$ whenever f_n is a sequence of morphisms converging to a morphism f.

The proof of Theorem 2 is given in Section 5. First we show how to deduce a characterization of Shannon entropy for general measures on finite sets.

The following definition is in analogy to Definition 1:

Definition 3. *Let* `FinMeas` *be the category whose objects are finite sets equipped with measures and whose morphisms are measure-preserving functions.*

There is more room for maneuver in `FinMeas` than in `FinProb`: we can take arbitrary nonnegative linear combinations of objects and morphisms, not just convex combinations. Any nonnegative linear combination can be built up from direct sums and multiplication by nonnegative scalars, which are defined as follows.

- For direct sums, first note that the disjoint union of two finite sets equipped with measures is another object of the same type. We write the disjoint union of $p, q \in$ `FinMeas` as $p \oplus q$. Then, given morphisms $f\colon p \to p'$, $g\colon q \to q'$ there is a unique morphism $f \oplus g\colon p \oplus q \to p' \oplus q'$ that restricts to f on the measure space p and to g on the measure space q.
- For scalar multiplication, first note that we can multiply a measure by a nonnegative real number and get a new measure. So, given an object $p \in$ `FinMeas` and a number $\lambda \geq 0$ we obtain an object $\lambda p \in$ `FinMeas` with the same underlying set and with $(\lambda p)_i = \lambda p_i$. Then, given a morphism $f\colon p \to q$, there is a unique morphism $\lambda f\colon \lambda p \to \lambda q$ that has the same underlying function as f.

This is consistent with our earlier notation for convex linear combinations.

We wish to give some conditions guaranteeing that a map sending morphisms in `FinMeas` to nonnegative real numbers comes from a multiple of Shannon entropy. To do this we need to define the Shannon entropy of a finite set X equipped with a measure p, not necessarily a probability measure. Define the **total mass** of (X, p) to be

$$\|p\| = \sum_{i \in X} p_i$$

If this is nonzero, then p is of the form $\|p\|\bar{p}$ for a unique probability measure space $\bar{p}$. In that case we define the **Shannon entropy** of p to be $\|p\|H(\bar{p})$. If the total mass of p is zero, we define its Shannon entropy to be zero.

We can define continuity for a map sending morphisms in `FinMeas` to numbers in $[0, \infty)$ just as we did for `FinProb`, and show:

Corollary 4. *Suppose F is any map sending morphisms in* `FinMeas` *to numbers in $[0, \infty)$ and obeying these four axioms:*

(i) ***Functoriality****:*

$$F(f \circ g) = F(f) + F(g)$$

whenever f, g are composable morphisms.

(ii) ***Additivity****:*

$$F(f \oplus g) = F(f) + F(g) \tag{3}$$

for all morphisms f, g.

(iii) ***Homogeneity****:*

$$F(\lambda f) = \lambda F(f) \tag{4}$$

for all morphisms f and all $\lambda \in [0, \infty)$.

(iv) Continuity: F is continuous.

Then there exists a constant $c \geq 0$ such that for any morphism $f \colon p \to q$ in `FinMeas`*,*

$$F(f) = c(H(p) - H(q))$$

where $H(p)$ is the Shannon entropy of p. Conversely, for any constant $c \geq 0$, this formula determines a map F obeying Conditions (i)–(iv).

Proof. Take a map F obeying these axioms. Then F restricts to a map on morphisms of `FinProb` obeying the axioms of Theorem 2. Hence there exists a constant $c \geq 0$ such that $F(f) = c(H(p) - H(q))$ whenever $f \colon p \to q$ is a morphism between probability measures. Now take an arbitrary morphism $f \colon p \to q$ in `FinMeas`. Since f is measure-preserving, $\|p\| = \|q\| = \lambda$, say. If $\lambda \neq 0$ then $p = \lambda \bar{p}$, $q = \lambda \bar{q}$ and $f = \lambda \bar{f}$ for some morphism $\bar{f} \colon \bar{p} \to \bar{q}$ in `FinProb`; then by homogeneity,

$$F(f) = \lambda F(\bar{f}) = \lambda c(H(\bar{p}) - H(\bar{q})) = c(H(p) - H(q)).$$

If $\lambda = 0$ then $f = 0f$, so $F(f) = 0$ by homogeneity. So $F(f) = c(H(p) - H(q))$ in either case. The converse statement follows from the converse in Theorem 2. □

3. Why Shannon Entropy Works

To prove the easy half of Theorem 2, we must check that $F(f) = c(H(p) - H(q))$ really does determine a functor obeying all the conditions of that theorem. Since all these conditions are linear in F, it suffices to consider the case where $c = 1$. It is clear that F is continuous, and Equation (1) is also immediate whenever $g \colon m \to p$, $f \colon p \to q$, are morphisms in `FinProb`:

$$F(f \circ g) = H(m) - H(q) = H(p) - H(q) + H(m) - H(p) = F(f) + F(g)$$

The work is to prove Equation (2).

We begin by establishing a useful formula for $F(f) = H(p) - H(q)$, where as usual f is a morphism $p \to q$ in `FinProb`. Since f is measure-preserving, we have

$$q_j = \sum_{i \in f^{-1}(j)} p_i$$

So

$$\begin{aligned} \sum_j q_j \ln q_j &= \sum_j \sum_{i \in f^{-1}(j)} p_i \ln q_j \\ &= \sum_j \sum_{i \in f^{-1}(j)} p_i \ln q_{f(i)} \\ &= \sum_i p_i \ln q_{f(i)} \end{aligned}$$

where in the last step we note that summing over all i that map to j and then summing over all j is the same as summing over all i. So,

$$\begin{aligned} F(f) &= -\sum_i p_i \ln p_i + \sum_j q_j \ln q_j \\ &= \sum_i (-p_i \ln p_i + p_i \ln q_{f(i)}) \end{aligned}$$

and thus

$$F(f) = \sum_{i \in X} p_i \ln \frac{q_{f(i)}}{p_i} \tag{5}$$

where the quantity in the sum is defined to be zero when $p_i = 0$. If we think of p and q as the distributions of random variables $x \in X$ and $y \in Y$ with $y = f(x)$, then $F(f)$ is exactly the conditional entropy of x given y. So, what we are calling "information loss" is a special case of conditional entropy.

This formulation makes it easy to check Equation (2),

$$F(\lambda f \oplus (1-\lambda) g) = \lambda F(f) + (1-\lambda) F(g)$$

simply by applying (5) on both sides.

In the proof of Corollary 4 (on `FinMeas`), the fact that $F(f) = c(H(p) - H(q))$ satisfies the four axioms was deduced from the analogous fact for `FinProb`. It can also be checked directly. For this it is helpful to note that

$$H(p) = \|p\| \ln \|p\| - \sum_i p_i \ln(p_i) \tag{6}$$

It can then be shown that Equation (5) holds for *every* morphism f in `FinMeas`. The additivity and homogeneity axioms follow easily.

4. Faddeev's Theorem

To prove the hard part of Theorem 2, we use a characterization of entropy given by Faddeev [4] and nicely summarized at the beginning of a paper by Rényi [5]. In order to state this result, it is convenient to write a probability measure on the set $\{1, \ldots, n\}$ as an n-tuple $p = (p_1, \ldots, p_n)$. With only mild cosmetic changes, Faddeev's original result states:

Theorem 5. ***(Faddeev)*** *Suppose I is a map sending any probability measure on any finite set to a nonnegative real number. Suppose that:*

(i) I is invariant under bijections.

(ii) I is continuous.

(iii) For any probability measure p on a set of the form $\{1, \ldots, n\}$, and any number $0 \leq t \leq 1$,

$$I((tp_1, (1-t)p_1, p_2, \ldots, p_n)) = I((p_1, \ldots, p_n)) + p_1 I((t, 1-t)) \tag{7}$$

Then I is a constant nonnegative multiple of Shannon entropy.

In Condition (i) we are using the fact that given a bijection $f\colon X \to X'$ between finite sets and a probability measure on X, there is a unique probability measure on X' such that p is measure-preserving; we demand that I takes the same value on both these probability measures. In Condition (ii), we use the standard topology on the simplex

$$\Delta^{n-1} = \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n \,\middle|\, p_i \geq 0,\ \sum_i p_i = 1 \right\}$$

to put a topology on the set of probability distributions on any n-element set.

The most interesting condition in Faddeev's theorem is (iii). It is known in the literature as the "grouping rule" ([6], Section 2.179)or "recursivity" ([2], Section 1.2.8). It is a special case of "strong additivity" ([2], Section 1.2.6), which already appears in the work of Shannon [1] and Faddeev [4]. Namely, suppose that p is a probability measure on the set $\{1, \ldots, n\}$. Suppose also that for each $i \in \{1, \ldots, n\}$, we have a probability measure $q(i)$ on a finite set X_i. Then $p_1 q(1) \oplus \cdots \oplus p_n q(n)$ is again a probability measure space, and the Shannon entropy of this space is given by the **strong additivity** formula:

$$H\Big(p_1 q(1) \oplus \cdots \oplus p_n q(n)\Big) = H(p) + \sum_{i=1}^{n} p_i H(q(i))$$

This can easily be verified using the definition of Shannon entropy and elementary properties of the logarithm. Moreover, Condition (iii) in Faddeev's theorem is equivalent to strong additivity together with the condition that $I((1)) = 0$, allowing us to reformulate Faddeev's theorem as follows:

Theorem 6. *Suppose I is a map sending any probability measure on any finite set to a nonnegative real number. Suppose that:*

(i) I is invariant under bijections.

(ii) I is continuous.

(iii) $I((1)) = 0$, where (1) is our name for the unique probability measure on the set $\{1\}$.

(iv) For any probability measure p on the set $\{1, \ldots, n\}$ and probability measures $q(1), \ldots, q(n)$ on finite sets, we have

$$I(p_1 q(1) \oplus \cdots \oplus p_n q(n)) = I(p) + \sum_{i=1}^{n} p_i I(q(i))$$

Then I is a constant nonnegative multiple of Shannon entropy. Conversely, any constant nonnegative multiple of Shannon entropy satisfies Conditions (i)–(iv).

Proof. Since we already know that the multiples of Shannon entropy have all these properties, we just need to check that Conditions (iii) and (iv) imply Faddeev's equation (7). Take $p = (p_1, \ldots, p_n)$, $q(1) = (t, 1-t)$ and $q(i) = (1)$ for $i \geq 2$: then Condition (iv) gives

$$I((tp_1, (1-t)p_1, p_2, \ldots, p_n)) = I((p_1, \ldots, p_n)) + p_1 I((t, 1-t)) + \sum_{i=2}^{n} p_i I((1))$$

which by Condition (iii) gives Faddeev's equation. □

It may seem miraculous how the formula

$$I(p_1, \ldots, p_n) = -c \sum_i p_i \ln p_i$$

emerges from the assumptions in either Faddeev's original Theorem 5 or the equivalent Theorem 6. We can demystify this by describing a key step in Faddeev's argument, as simplified by Rényi [5]. Suppose I is a function satisfying the assumptions of Faddeev's result. Let

$$\phi(n) = I\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$$

equal I applied to the uniform probability measure on an n-element set. Since we can write a set with nm elements as a disjoint union of m different n-element sets, Condition (iv) of Theorem 6 implies that

$$\phi(nm) = \phi(n) + \phi(m)$$

The conditions of Faddeev's theorem also imply

$$\lim_{n \to \infty} (\phi(n+1) - \phi(n)) = 0$$

and the only solutions of both these equations are

$$\phi(n) = c \ln n$$

This is how the logarithm function enters. Using Condition (iii) of Theorem 5, or equivalently Conditions (iii) and (iv) of Theorem 6, the value of I can be deduced for probability measures p such that each p_i is rational. The result for arbitrary probability measures follows by continuity.

5. Proof of the Main Result

Now we complete the proof of Theorem 2. Assume that F obeys Conditions (i)–(iii) in the statement of this theorem.

Recall that (1) denotes the set $\{1\}$ equipped with its unique probability measure. For each object $p \in \mathtt{FinProb}$, there is a unique morphism

$$!_p \colon p \to (1)$$

We can think of this as the map that crushes p down to a point and loses all the information that p had. So, we define the "entropy" of the measure p by

$$I(p) = F(!_p)$$

Given any morphism $f\colon p \to q$ in `FinProb`, we have

$$!_p = !_q \circ f$$

So, by our assumption that F is functorial,

$$F(!_p) = F(!_q) + F(f)$$

or in other words:

$$F(f) = I(p) - I(q) \tag{8}$$

To conclude the proof, it suffices to show that I is a multiple of Shannon entropy.

We do this by using Theorem 6. Functoriality implies that when a morphism f is invertible, $F(f) = 0$. Together with (8), this gives Condition (i) of Theorem 6. Since $!_{(1)}$ is invertible, it also gives Condition (iii). Condition (ii) is immediate. The real work is checking Condition (iv).

Given a probability measure p on $\{1, \ldots, n\}$ together with probability measures $q(1), \ldots, q(n)$ on finite sets $X_1, \ldots, X_n$, respectively, we obtain a probability measure $\bigoplus_i p_i q(i)$ on the disjoint union of $X_1, \ldots, X_n$. We can also decompose p as a direct sum:

$$p \cong \bigoplus_i p_i(1) \tag{9}$$

Define a morphism

$$f = \bigoplus_i p_i !_{q(i)} \colon \bigoplus_i p_i q(i) \to \bigoplus_i p_i(1)$$

Then by convex linearity and the definition of I,

$$F(f) = \sum_i p_i F(!_{q(i)}) = \sum_i p_i I(q(i))$$

But also

$$F(f) = I\big(\bigoplus_i p_i q(i)\big) - I\big(\bigoplus p_i(1)\big) = I\big(\bigoplus_i p_i q(i)\big) - I(p)$$

by (8) and (9). Comparing these two expressions for $F(f)$ gives Condition (iv) of Theorem 6, which completes the proof of Theorem 2.

6. A Characterization of Tsallis Entropy

Since Shannon defined his entropy in 1948, it has been generalized in many ways. Our Theorem 2 can easily be extended to characterize one family of generalizations, the so-called "Tsallis entropies". For any positive real number α, the **Tsallis entropy of order** α of a probability measure p on a finite set X is defined as:

$$H_\alpha(p) = \begin{cases} \dfrac{1}{\alpha - 1}\Big(1 - \displaystyle\sum_{i \in X} p_i^\alpha\Big) & \text{if } \alpha \neq 1 \\ -\displaystyle\sum_{i \in X} p_i \ln p_i & \text{if } \alpha = 1 \end{cases}$$

The peculiarly different definition when $\alpha = 1$ is explained by the fact that the limit $\lim_{\alpha \to 1} H_\alpha(p)$ exists and equals the Shannon entropy $H(p)$.

Although these entropies are most often named after Tsallis [7], they and related quantities had been studied by others long before the 1988 paper in which Tsallis first wrote about them. For example, Havrda and Charvát [8] had already introduced a similar formula, adapted to base 2 logarithms, in a 1967 paper in information theory, and in 1982, Patil and Taillie [9] had used H_α itself as a measure of biological diversity.

The characterization of Tsallis entropy is exactly the same as that of Shannon entropy except in one respect: in the convex linearity condition, the degree of homogeneity changes from 1 to α.

Theorem 7. *Let $\alpha \in (0, \infty)$. Suppose F is any map sending morphisms in* `FinProb` *to numbers in $[0, \infty)$ and obeying these three axioms:*

(i) ***Functoriality****:*

$$F(f \circ g) = F(f) + F(g)$$

whenever f, g are composable morphisms.

(ii) ***Compatibility with convex combinations****:*

$$F(\lambda f \oplus (1 - \lambda) g) = \lambda^\alpha F(f) + (1 - \lambda)^\alpha F(g)$$

for all morphisms f, g and all $\lambda \in [0, 1]$.

(iii) ***Continuity****: F is continuous.*

Then there exists a constant $c \geq 0$ such that for any morphism $f \colon p \to q$ in `FinProb`*,*

$$F(f) = c(H_\alpha(p) - H_\alpha(q))$$

where $H_\alpha(p)$ is the order α Tsallis entropy of p. Conversely, for any constant $c \geq 0$, this formula determines a map F obeying Conditions (i)–(iii).

Proof. We use Theorem V.2 of Furuichi [10]. The statement of Furuichi's theorem is the same as that of Theorem 5 (Faddeev's theorem), except that Condition (iii) is replaced by

$$I((tp_1, (1-t)p_1, p_2, \ldots, p_n)) = I((p_1, \ldots, p_n)) + p_1^\alpha I((t, 1-t))$$

and Shannon entropy is replaced by Tsallis entropy of order α. The proof of the present theorem is thus the same as that of Theorem 2, except that Faddeev's theorem is replaced by Furuichi's. □

As in the case of Shannon entropy, this result can be extended to arbitrary measures on finite sets. For this we need to define the Tsallis entropies of an arbitrary measure on a finite set. We do so by requiring that

$$H_\alpha(\lambda p) = \lambda^\alpha H_\alpha(p)$$

for all $\lambda \in [0, \infty)$ and all $p \in$ `FinMeas`. When $\alpha = 1$ this is the same as the Shannon entropy, and when $\alpha \neq 1$, we have

$$H_\alpha(p) = \frac{1}{\alpha - 1} \left(\left(\sum_{i \in X} p_i \right)^\alpha - \sum_{i \in X} p_i^\alpha \right)$$

(which is analogous to (6)). The following result is the same as Corollary 4 except that, again, the degree of homogeneity changes from 1 to α.

Corollary 8. *Let $\alpha \in (0, \infty)$. Suppose F is any map sending morphisms in* `FinMeas` *to numbers in $[0, \infty)$, and obeying these four properties:*

(i) ***Functoriality****:*

$$F(f \circ g) = F(f) + F(g)$$

whenever f, g are composable morphisms.

(ii) ***Additivity****:*

$$F(f \oplus g) = F(f) + F(g)$$

for all morphisms f, g.

(iii) ***Homogeneity of degree*** α*:*

$$F(\lambda f) = \lambda^\alpha F(f)$$

for all morphisms f and all $\lambda \in [0, \infty)$.

(iv) ***Continuity****: F is continuous.*

Then there exists a constant $c \geq 0$ such that for any morphism $f \colon p \to q$ in `FinMeas`*,*

$$F(f) = c(H_\alpha(p) - H_\alpha(q))$$

where H_α is the Tsallis entropy of order α. Conversely, for any constant $c \geq 0$, this formula determines a map F obeying Conditions (i)–(iv).

Proof. This follows from Theorem 7 in just the same way that Corollary 4 follows from Theorem 2. □

Acknowledgements

We thank the denizens of the n-Category Café, especially David Corfield, Steve Lack, Mark Meckes and Josh Shadlen, for encouragement and helpful suggestions. Tobias Fritz is supported by the EU STREP QCS. Tom Leinster is supported by an EPSRC Advanced Research Fellowship.

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
2. Aczél, J.; Daróczy, Z. On Measures of information and their characterizations. In *Mathematics in Science and Engineering*; Academic Press: New York, NY, USA, 1975; Volume 15.
3. Mac Lane, S. Categories for the working mathematician. In *Graduate Texts in Mathematics 5*; Springer: Berlin, Germany, 1971.
4. Faddeev, D.K. On the concept of entropy of a finite probabilistic scheme. *Uspehi Mat. Nauk (N.S.)* **1956**, *1*, 227–231. (In Russian).

5. Rényi, A. Measures of information and entropy. In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Statistical Laboratory of the University of California, Berkeley, CA, USA, June 20–July 30, 1960; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
6. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley Series in Telecommunications and Signal Processing; Wiley-Interscience: New York, NY, USA, 2006.
7. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
8. Havrda, J.; Charvát, F. Quantification method of classification processes: concept of structural α-entropy. *Kybernetika* **1967**, *3*, 30–35.
9. Patil, G.P.; Taillie, C. Diversity as a concept and its measurement. *J. Am. Stat. Assoc.* **1982**, *77*, 548–561.
10. Furuichi, S. On uniqueness theorems for Tsallis entropy and Tsallis relative entropy, *IEEE Trans. Infor. Theor.* **2005**, *51*, 3638–3645.