Which Bio-Diversity Indices Are Most Adequate

Olga Kosheleva
Department of Teacher Education
University of Texas at El Paso
500 W. University
El Paso, Texas 79968
olgak@utep.edu

Craig Tweedie
Environmental Science and Engineering Program
University of Texas at El Paso
500 W. University
El Paso, Texas 79968
ctweedie@utep.edu

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968
vladik@utep.edu

***Abstract*—One of the main objectives of ecology is to analyze, maintain, and enhance the bio-diversity of different ecosystems. To be able to do that, we need to gauge bio-diversity. Several semi-heuristic diversity indices have been shown to be in good accordance with the intuitive notion of bio-diversity. In this paper, we provide a theoretical justification for these empirically successful techniques. Specifically, we show that the most widely used techniques – Simpson index – can be justified by using simple fuzzy rules, while a more elaborate justification explains all empirically successful diversity indices.**

I. INTRODUCTION

Gauging bio-diversity is important. One of the main objectives of ecology is to study and preserve bio-diversity. To compare different situations and different strategies, it is therefore important to have a numerical measure of bio-diversity that would adequately describe the diversity in a given area.

How bio-diversity is measured now. Most existing measures of diversity are based on the relative frequencies p_i of different species. The most widely used measures are the *Shannon index* [14]

$$H = -\sum_{i=1}^{n} p_i \cdot \ln(p_i), \tag{1}$$

and the *Simpson index* [15]

$$D = \sum_{i=1}^{n} p_i^2 \tag{2}$$

(see, e.g., [1], [2]). The Simpson index is also known as the *Herfindahl index* [6] or the *Herfindal-Hirschman index* (HHI) [7].

Both measures attain their largest values when there are no species is rare than others, i.e., when when all the frequencies are the same

$$p_1 = \ldots = p_n = \frac{1}{n}.$$

In addition to the original Simpson index D, ecologists also use indices which are related to D such as the *inverse Simpson index*

$$\frac{1}{D}$$

and the *Gini-Simpson index* $1 - D$. The Gini-Simpson index is also known as the *probability of interspecific encounter* (PIE), as the *Gibbs-Martin index* [5], and as the *Blau index.*

In addition to the Simpson index, which is the sum of squares of corresponding frequencies, ecologists also use indices related to the sum of the q-th powers

$$\sum_{i=1}^{n} p_i^q, \tag{3}$$

for an arbitrary real value q. For example, they use *Rényi entropy* [13]

$$H_q = \frac{1}{1-q} \cdot \ln\left(\sum_{i=1}^{n} p_i^q\right). \tag{4}$$

Rënyi entropy if not directly defined for $q = 1$, but we can defined if we take a limit $q \to 1$. In this limit, the expression (4) tends to the Shannon index (1).

Why these measures? Experience has shown that the above measures of diversity are, empirically, in good accordance with the ecologists' intuition. However, from the theoretical viewpoint, the success of these particular measures of diversity is somewhat puzzling: why these expressions and not other possible expressions?

What we do in this paper. In this paper, we provide possible justification for the above empirically successful measures. To be more precise, we provide two possible justification:

- we start with a simple fuzzy logic-based justification which explains Simpson index, and then
- we provide a more elaborate justification that explains all the above diversity measures.

Comment. Our "more elaborate" justification uses mathematical techniques similar to the ones that we used in [9] for a different practical problem: how to select the most appropriate image reconstruction technique.

II. A STRAIGHTFORWARD FUZZY-BASED EXPLANATION OF SIMPSON INDEX

An intuitive meaning of bio-diversity. An ecosystem is perfectly diverse if all the species that form this ecosystem are reasonably frequent but not dominant. In other words, the ecosystem is healthy if:

- the first species is reasonably frequent but not dominant, and

- the second species is reasonably frequent but not dominant,
- etc.

Let us translate this intuitive meaning into a precise measure of diversity. The above intuitive statement is not precise, since it uses an imprecise ("'fuzzy") natural-language terms like "reasonably frequent" and "not dominant". We need to translate this statement into precise terms.

The need to translate statements containing imprecise ("fuzzy") words from natural language into precise terms was recognized in the early 1960s by L. Zadeh who invented, for such translations, a special techniques that he called *fuzzy logic*; see, e.g., [8], [12], [17]. Let us therefore use fuzzy logic techniques to translate the above statement into precise terms.

What is the meaning of "reasonably frequent". The above statement consists of several elementary statements combined by "and". In accordance with the general fuzzy logic methodology, let us first describe each elementary statement in precise terms. For that, we need to assign a precise meaning to the terms "reasonably frequent" and "not dominant".

In fuzzy logic, the meaning of each imprecise word is described by a *membership function* that assigns, to each possible value x of the corresponding quantity, a degree $\mu(x)$ to which this value satisfies the corresponding imprecise property. In our case, we thus need to assign, to each possible value of frequency p_i, a degree $\mu(p_i)$ to which the corresponding species is reasonably frequent and not dominant.

Let us combine the elementary statements into a single composite statement describing diversity. We are interested in the degree to which a given ecosystem is diverse. As we have mentioned, the ecosystem is diverse if the above elementary statement holds for the first species *and* for the second species, etc.

We already know the degree $\mu(p_i)$ to which each individual statement holds. In fuzzy logic, the degree to which an "and"-combination of several statements is true, we apply a special "and"-operation $f_\& : [0,1] \times [0,1] \to [0,1]$ (also known as a *t-norm*) to the degrees to which individual statements are true.

The general strategy in applications of fuzzy techniques is to select the simplest possible "and"-operation; see, e.g., [8], [12]. In line with this idea, the two most frequently used (and more practically successful) "and"-operations are the product and the minimum.

Minimum is somewhat easier to compute, but let us recall that our objective is not simply to gauge diversity, but to come up with some recommendations for boosting bio-diversity. In other words, our objective is not so much to estimate, but rather to use the resulting estimates as an objective function whose value we try to maximize.

Most efficient optimization techniques use differentiation. Actually, optimization was one of the main reasons why calculus was invented in the first place – since it turned out that to find the optimum of a function, it is sufficient to find the point where its derivative is equal to 0. From this viewpoint, it is desirable to come up with the differentiable measure of diversity.

This desire eliminates minimum, since $\min(a,b)$ is not always differentiable: namely, it is not differentiable when $a = b$. Thus, of the two simplest "and"-operations, we should select the product.

What is the resulting measure of diversity. The resulting measure of diversity is thus equal to the product

$$\prod_{i=1}^{n} \mu(p_i) \tag{5}$$

of the values $\mu(p_i)$ corresponding to all the species from this ecosystem.

This expression can be somewhat simplified if we take into account that maximizing the product is equivalent to maximizing its logarithm L, and that the logarithm of the product is equal to the sum of the logarithms. Thus, maximizing the product is equivalent to maximizing the expression

$$L = \sum_{i=1}^{n} f(p_i), \tag{6}$$

where we denoted $f(p_i) \stackrel{\text{def}}{=} \ln(\mu(p_i))$.

In a diverse ecosystem all the frequencies p_i are rather small – if one of the values is large, this means that we have a dominant species, the ecosystem is no longer diverse. For small p_i, we can replace each value $f(p_i)$ with the sum of the few first terms in its Taylor expansion.

In the first approximation, we can take $f(p_i) = a_0 + a_1 \cdot p_i$, and thus,

$$L = \sum_{i=1}^{n}(a_0 + a_1 \cdot p_i) = \sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 \cdot p_i = a_0 \cdot n + a_1 \cdot \sum_{i=1}^{n} p_i.$$

The sum of all the frequencies p_i is 1, so in this approximation, $L = a_0 \cdot n + a_1$. This expression does not depend on the frequencies p_i at all and thus, cannot serve as a reasonable measure of diversity.

So, to adequately describe diversity, we need to take into account quadratic terms in the Taylor expansion, i.e., take $f(p_i) = a_0 + a_1 \cdot p_i + a_2 \cdot p_i^2$. In this approximation,

$$L = \sum_{i=1}^{n}(a_0 + a_1 \cdot p_i + a_2 \cdot p_i^2) = a_0 \cdot n + a_1 + a_2 \cdot \sum_{i=1}^{n} p_i^2.$$

Since a_0, a_1, and a_2 are constants, maximizing this expression is equivalent to maximizing the sum

$$\sum_{i=1}^{n} p_i^2.$$

So, we have indeed justified the use of Simpson index.

III. A MORE ELABORATE JUSTIFICATION THAT EXPLAINS ALL EMPIRICALLY SUCCESSFUL DIVERSITY MEASURES

The ultimate purpose of diversity estimation is decision making. As we have mentioned, the ultimate purpose of gauging uncertainty is to make a decision. When we have a diverse ecosystem, we need to decide how to preserve and maintain the corresponding healthy level of bio-diversity. In situations when the level of diversity is far from ideal, we need to come up with some recommendations on how to improve bio-diversity.

From this viewpoint, what we really want is to describe the expert's preferences: which combinations of frequencies $p = (p_1, \ldots, p_n)$ are preferred and which are not.

How to describe preferences? There exists a general formalism. The necessity to describe *preferences* is extremely important in decision making in general, including decision making under conflict (also known under a somewhat misleading name of *game theory*). To describe these preferences, a special *utility theory* has been developed; see, e.g., [3], [10], [11], [16].

The mathematical formalism of utility theory comes from the observation that sometimes, when a person faces several alternatives $A_1, \ldots, A_n$, instead of choosing one of these alternatives, this person may choose a *probabilistic* combination of them, i.e., A_1 with probability P_1, A_2 with a probability P_2, etc. For example, if two alternatives are of equal value to a person, that person will probably choose the first one with probability 0.5 and the second one with the same probability 0.5. Such probabilistic combinations are called (somewhat misleadingly) *lotteries*.

For example, in ecology, in addition to (often very costly) plans that guarantee that bio-diversity improves, it makes sense to also consider more affordable plans that improve bio-diversity only with a certain probability.

In view of this realistic possibility, it is desirable to consider the preference relation not only for the original alternatives, but also for arbitrary lotteries combining these alternatives. Each original alternative A_i can be viewed as a *degenerate* lottery, in which this alternative A_i appears with probability 1, and every other alternative $A_j \neq A_i$ appear with probability 0.

The main result of utility theory states that if we have an ordering relation $L \succeq L'$ between such lotteries (with the meaning "L is preferable to L'"), and if this relation satisfies natural consistency conditions such as transitivity, etc., then there exists a function u from the set $\mathcal{L}$ of all possible lotteries into the set R of real numbers for which:

- $L \succeq L'$ if and only if $u(L) \geq u(L')$, and
- for every lottery L, in which each alternative A_i appears with probability p_i, we have

$$u(L) = P_1 \cdot u(A_1) + \ldots + P_n \cdot u(A_n).$$

This function u is called a *utility function*. Each consistent preference relation can thus be described by its utility function.

In our case, to describe preferences between frequency tuples $p = (p_1, \ldots, p_n)$, we need a utility function $u(p_1, \ldots, p_n)$ that is defined on the set of all possible tuples.

Which function should we choose?

Localness property: intuitive idea. An important intuitive feature of bio-diversity is the *localness* property. This property is motivated by the fact that, in addition to the bio-diversity of the whole ecosystem, we may be interested in the bio-diversity of its subsystem.

For the whole ecosystem, the sum of frequencies is 1. When we analyze a subsystem, we only take into account some of the species, so the sum of the frequencies can be smaller than 1.Thus, we need to consider the values $u(p)$ for tuples for which $\sum\limits_i p_i < 1$.

It should be mentioned, however, that while it makes sense to compare two possible arrangement within the ecosystem as a whole, or two possible arrangements within a subsystem of this ecosystem, we do not consider a more complex problem of comparing the bio-diversity of the whole ecosystem with the bio-diversity of a subsystem. In precise terms, it means that only compare tuples $p = (p_1, \ldots, p_n)$ and $p' = (p'_1, \ldots, p'_n)$ for which

$$\sum_{i=1}^n p_i = \sum_{i=1}^n p'_i.$$

Let us now consider two tuples $p = (p_1, \ldots, p_n)$ and $p' = (p'_1, \ldots, p'_n)$ which are, in this sense, comparable. Let us assume that for some species, the frequencies are the same, i.e., $p_i = p'_i$ for all indices i from some set I. Suppose also that, from the point of bio-diversity, the tuple p is preferable to tuple p': $p \succeq p'$.

Intuitively, this means that while in the two tuples, the level of diversity is the same for species from the set I, species from the complement set $-I$ have a higher degree of bio-diversity. Thus, if we replace the values $p_i = p'_i$ corresponding to species $i \in I$ with some other values $q_i = q'_i$ while preserving the same frequencies for species $i \notin I$ (i.e., $q_i = p_i$ and $q'_i = p'_i$ for such i), then the resulting tuple q will still correspond to a higher degree of bio-diversity the tuple q': $q \succeq q'$.

Localness property: description in precise terms. In precise terms, this localness property takes the following form. Let $I \subseteq \{1, \ldots, n\}$ be a set of indices, let $p \succeq p'$ be two tuples for which $p_i = p'_i$ for all $i \in I$, and let q and q' be another two tuples for which:

- $q_i = p_i$ and $q'_i = p'_i$ for all $i \notin I$; and
- $q_i = q'_i$ for all $i \in I$.

Then, $q \succeq q'$.

Consequences of localness property. Such a localness property (also known as *independence property*) is a frequent feature in practical problems, and utility theory has developed a precise description of utility functions that satisfy this property. Namely, it has been shown that when alternatives are characterized by n parameters $x_1, \ldots, x_n$, then the localness of the preference is equivalent to the utility function $u(x_1, \ldots, x_n)$ being of one of the two types [4]:

- *additive* $u(x_1, \ldots, x_n) = u_1(x_1) + \ldots + u_n(x_n)$ for some functions $u_i(x_i)$; or
- *multiplicative* $U(x_1, \ldots, x_n) = U_1(x_1) \cdot \ldots \cdot U_n(x_n)$ for some functions $U_i(x_i)$.

Multiplicative case can be reduced to an additive one. Our objective is to compare different tuples. We are not interested in specific values of utility, we only want to find out which tuples correspond to higher degree of bio-diversity. Since logarithm is a strictly increasing function, the comparison $U(p) \geq U(p')$ is equivalent to $\ln(U(p)) \geq \ln(U(p'))$. So, instead of the original degree $U(p)$, we can also use $\ln(U(p))$ as a measure of bio-diversity.

For the product utility

$$U(p_1, \ldots, p_n) = \prod_{i=1}^{n} U_i(p_i),$$

its logarithm is the sum of the corresponding logarithms

$$\ln(U(p)) = \sum_{i=1}^{n} \ln(U_i(p_i)).$$

Thus, in the multiplicative case, we can use an additive measure of bio-diversity

$$u(p) = \sum_{i=1}^{n} u_i(p_i),$$

where $u(p) \stackrel{\text{def}}{=} \ln(U(p))$ and $u_i(p_i) \stackrel{\text{def}}{=} \ln(u_i(p_i))$.

Thus, in both case, we can use an additive measure of bio-diversity

$$u(p) = \sum_{i=1}^{n} u_i(p_i).$$

The degree of bio-diversity should not change if we change the arbitrary numbers assigned to different species. In our formulation of the problem, we only take into account the frequencies of different species. In this formulation, numbers assigned to species – which species is number 1, which is number 2, etc. – are arbitrary. So, if we simply change these arbitrarily selected numbers, the degree of bio-diversity should not change.

Thus, the dependence of u_i on p_i should not depend on i, i.e., we should have $u_i(p_i) = d(p_i)$ for one of the same function $d(p)$. In this case, the desired degree of bio-diversity is equal to

$$u(p) = \sum_{i=1}^{n} d(p_i).$$

So, the question is which functions $d(p)$ are appropriate for describing bio-diversity.

Without losing generality, we can assume that the function $d(p)$ is twice differentiable. Since our ultimate goal is optimization, it is desirable to consider only *smooth* (differentiable) functions $d(p)$, because for smooth functions, optimization is as easy as computing the derivatives and equating them to 0. Moreover, since many useful optimization techniques use the second derivatives as well, it is desirable to consider only *twice* differentiable functions.

Fortunately, we can impose this restriction without losing generality, because, as it is well known, every continuous function can be, with an arbitrary accuracy, approximated by twice differentiable functions (even by polynomials). Since we are dealing with not 100% accurate data anyway, there is no reason to compute the bio-diversity measure absolutely precisely. Therefore, even if the actual expert opinion on bio-diversity is best described by a non-smooth function $d(p)$, we can, within any given accuracy, still approximate it by a smooth function. Because of this possibility, in the following text, we will assume that the desired function $d(p)$ is twice differentiable.

Possibility of scaling. We have mentioned earlier that we can either consider bio-diversity of the ecosystem as a whole, or, alternatively, bio-diversity of its subsystem. Strictly speaking, however, this distinction is rather artificial: in nature, everything is related, so each ecosystem can be viewed as a subsystem of a larger ecosystem – all the way to the entire biosphere.

Whether we consider an ecosystem by itself or as a part of larger ecosystem changes the corresponding frequencies. If we consider an ecosystem by itself, then the corresponding frequencies add up to 1:

$$\sum_{i=1}^{n} p_i = 1.$$

Alternatively, when we consider this ecosystem as a part of the larger ecosystem, the frequencies change: previously, each frequency was determined as

$$p_i = \frac{n_i}{n},$$

where n_i is the total population of the i-th species in the given area and n is the total bio-population in this area. Now, instead, we have the ratio

$$\frac{n_i}{N},$$

where N is the total bio-population of the larger area. Thus, the new frequencies have the form $\lambda \cdot p_i$, where we denoted

$$\lambda \stackrel{\text{def}}{=} \frac{n}{N}.$$

So, instead of the original tuple $p = (p_1, \ldots, p_n)$, we have a new tuple $\lambda \cdot p \stackrel{\text{def}}{=} (\lambda \cdot p_1, \ldots, \lambda \cdot p_n)$.

Relative bio-diversity of a region should not depend on whether we consider this region as a separate ecosystem or as a part of a larger ecosystem. Thus, if we have $p \succeq p'$ for two tuples, we should also have $\lambda \cdot p \succeq \lambda \cdot p'$.

Since $p \succeq p'$ is equivalent to

$$\sum_{i=1}^{n} d(p_i) \geq \sum_{i=1}^{n} d(p'_i),$$

we thus arrive at the following precise description of the scaling property.

Definition 1. *We say that a twice differentiable function $d(p)$ is* scale-invariant *if for every two tuples $p = (p_1, \ldots, p_n)$ and $p' = (p'_1, \ldots, p'_n)$ for which*

$$\sum_{i=1}^n p_i = \sum_{i=1}^n p'_i \text{ and } \sum_{i=1}^n d(p_i) = \sum_{i=1}^n d(p'_i),$$

and for every real number $\lambda > 0$, we have

$$\sum_{i=1}^n d(\lambda \cdot p_i) = \sum_{i=1}^n d(\lambda \cdot p'_i).$$

Definition 2. *We say that two functions $d_1(p)$ and $d_1(p)$ are* equivalent *if $d_2(p) = a + b \cdot p + c \cdot d_1(p)$ for some constants a and $b > 0$.*

Comment. When the functions $d_1(p)$ and $d_2(p)$ are equivalent, then

$$\sum_{i=1}^n d_2(p_i) = a \cdot n + b + c \cdot \sum_{i=1}^n d_1(p_i).$$

Thus, when we compare different tuples, these two functions $d_1(p)$ and $d_2(p)$ lead to the same conclusions on which tuple corresponds to larger bio-diversity.

Proposition. *Every scale-invariant function is equivalent to $d(p) = \pm \ln(p)$, $d(p) = \pm p^q$ for some q, or to*

$$d(p) = \pm p \cdot \ln(p).$$

Discussion. The corresponding sums

$$\sum_{i=1}^n d(p_i)$$

are exactly Shannon, Simpson, and Rényi indices. Thus, we have indeed explained why exactly these indices adequately describe expert's intuition about bio-diversity.

Proof.

1°. Let us consider small deviations $p'_i = p_i + \varepsilon \cdot \Delta p_i$ for some small ε. Then,

$$\sum_{i=1}^n p'_i = \sum_{i=1}^n (p_i + \varepsilon \cdot \Delta p_i) = \sum_{i=1}^n p_i + \varepsilon \cdot \sum_{i=1}^n \Delta p_i;$$

$$\sum_{i=1}^n d(p'_i) = \sum_{i=1}^n d(p_i + \varepsilon \cdot \Delta p_i) =$$

$$\sum_{i=1}^n (d(p_i) + d'(p_i) \cdot \varepsilon \cdot \Delta p_i + O(\varepsilon^2)) =$$

$$\sum_{i=1}^n d(p_i) + \varepsilon \cdot \sum_{i=1}^n d'(p_i) \cdot \Delta p_i + O(\varepsilon^2);$$

and

$$\sum_{i=1}^n d(\lambda \cdot p'_i) \approx \sum_{i=1}^n d(\lambda \cdot p_i + \lambda \cdot \varepsilon \cdot \Delta p_i) =$$

$$\sum_{i=1}^n (d(\lambda \cdot p_i) + d'(\lambda \cdot p_i) \cdot \lambda \cdot \varepsilon \cdot \Delta p_i + O(\varepsilon^2)) =$$

$$\sum_{i=1}^n d(\lambda \cdot p_i) + \lambda \cdot \varepsilon \cdot \sum_{i=1}^n d'(\lambda \cdot p_i) \cdot \Delta p_i + O(\varepsilon^2),$$

where $d'(p)$ denotes the derivative of the function $d(p)$.

Thus, for these pairs p and p', scale-invariance means that if

$$\sum_{i=1}^n \Delta p_i = 0 \text{ and } \sum_{i=1}^n d'(p_i) \cdot \Delta p_i + O(\varepsilon) = 0,$$

then

$$\sum_{i=1}^n d'(\lambda \cdot p_i) \cdot \Delta p_i + O(\varepsilon) = 0.$$

In the limit $\varepsilon \to 0$, we have the following implication: if

$$\sum_{i=1}^n \Delta p_i = 0 \text{ and } \sum_{i=1}^n d'(p_i) \cdot \Delta p_i = 0,$$

then

$$\sum_{i=1}^n d'(\lambda \cdot p_i) \cdot \Delta p_i = 0.$$

2°. The above property can be reformulated in geometric terms, if we take into account that the sum

$$\sum_{i=1}^n \Delta p_i$$

is a dot (scalar) product between the vector $\Delta p = (\Delta p_1, \ldots, \Delta p_n)$ and a unit vector $e \stackrel{\text{def}}{=} (1, \ldots, 1)$:

$$\sum_{i=1}^n \Delta p_i = \Delta p \cdot e.$$

In this terms, the fact that $\Delta p \cdot e = 0$ means that the vector Δp is orthogonal to the vector e.

Similarly,

$$\sum_{i=1}^n d'(p_i) \cdot \Delta p_i = d' \cdot \Delta p,$$

where $d' \stackrel{\text{def}}{=} (d'(p_1), \ldots, d'(p_n))$, and

$$\sum_{i=1}^n d'(\lambda \cdot p_i) \cdot \Delta p_i = d'_\lambda \cdot \Delta p,$$

where $d'_\lambda \stackrel{\text{def}}{=} (d'(\lambda \cdot p_1), \ldots, d'(\lambda \cdot p_n))$.

Thus, the above property says that any vector Δp which is orthogonal to both vectors e and d' is also orthogonal to the vector d'_λ.

In geometric terms, it is easy to prove that if a vector v is orthogonal to every vector x that is orthogonal to two given vectors v_1 and v_2, then v belongs to the linear space generated by v_1 and v_2: indeed, otherwise, we could take a projection $\pi(v)$ of v on the orthogonal complement to that linear space;

this projection is orthogonal to both v_i, but not to v. Thus, for every tuple p and for every λ, we have

$$d'_\lambda = \alpha(\lambda, p) \cdot e + \beta(\lambda, p) \cdot d'$$

for some coefficients α and β which, in general, depend on λ and on p. In terms of the vector components, this means that for every i, we have

$$d'(\lambda \cdot p_i) = \alpha(\lambda, p) + \beta(\lambda, p) \cdot d'(p_i). \tag{7}$$

3°. Let us show that the values α and β depend only on λ and do not depend on the tuple p. Let us first prove this for β. Indeed, if we subtract the equations (7) corresponding to two different indices i and j, we conclude that

$$d'(\lambda \cdot p_i) - d'(\lambda \cdot p_j) = \beta(\lambda, p) \cdot (d'(p_i) - d'(p_j)), \tag{8}$$

i.e., that

$$\beta(\lambda, p) = \frac{d'(\lambda \cdot p_i) - d'(\lambda \cdot p_j)}{d'(p_i) - d'(p_j)}. \tag{9}$$

The right-hand side of this equality only depends on p_i and p_j and does not depend on any other frequencies p_k. Thus, the coefficient $\beta(\lambda, p)$ only depends on p_i and p_j and does not depend on any other frequencies p_k.

If we consider a similar formula with a different pair (i', j'), we will conclude that $\beta(\lambda, p)$ does not depend on the frequencies p_i and p_j either. Thus, β does not depend on the tuple p at all, it only depends on λ: $\beta(\lambda, p) = \beta(\lambda)$.

Thus, the formula (7) has the form

$$d'(\lambda \cdot p_i) = \alpha(\lambda, p) + \beta(\lambda) \cdot d'(p_i), \tag{10}$$

hence

$$\alpha(\lambda, p) = d'(\lambda \cdot p_i) - \beta(\lambda) \cdot d'(p_i). \tag{11}$$

The right-hand side of this formula only depends on p_i and does not depend on any other frequency p_j. Thus, the coefficient $\alpha(\lambda, p)$ only depends on p_i and does not depend on any other frequency p_j.

If we consider a similar formula with a different index i', we will conclude that $\alpha(\lambda, p)$ does not depend on the frequency p_i either. Thus, α does not depend on the tuple p at all, it only depends on λ: $\alpha(\lambda, p) = \alpha(\lambda)$. So, the formulas (7) and (10) take the form

$$d'(\lambda \cdot p_i) = \alpha(\lambda) + \beta(\lambda) \cdot d'(p_i). \tag{12}$$

4°. We assumed that the function $d(p)$ is twice differentiable. Thus, the function $D(p) \stackrel{\text{def}}{=} d'(p)$ is differentiable. In terms of this function, the equality (12) take the form

$$D(\lambda \cdot p_i) = \alpha(\lambda) + \beta(\lambda) \cdot D(p_i). \tag{13}$$

Let us prove that the functions $\alpha(\lambda)$ and $\beta(\lambda)$ are also differentiable. Indeed, from the formula (9), we conclude that

$$\beta(\lambda) = \frac{D(\lambda \cdot p_i) - D(\lambda \cdot p_j)}{D(p_i) - D(p_j)}. \tag{14}$$

Since the function $D(p)$ is differentiable, the right-hand side of the formula (14) is differentiable and thus, its left-hand side (i.e., the function $\beta(\lambda)$) is differentiable as well.

From the formula (11), it follows that

$$\alpha(\lambda) = D(\lambda \cdot p_i) - \beta(\lambda) \cdot D(p_i). \tag{15}$$

Since the functions $D(p)$ and $\beta(\lambda)$ are differentiable, we can thus conclude that the function $\alpha(\lambda)$ is differentiable as well.

5°. Now, we are ready to deduce the differential equation from the functional equation (13).

Since all three functions $D(p)$, $\alpha(\lambda)$, and $\beta(\lambda)$, are differentiable, we can differentiate both sides of the equation (13) with respect to λ and substitute $\lambda = 1$. As a result, we get the following differential equation:

$$p \cdot \frac{dD}{dp} = A + B \cdot D, \tag{16}$$

where we denoted $A \stackrel{\text{def}}{=} \alpha'(1)$ and $B \stackrel{\text{def}}{=} \beta'(1)$.

6°. To solve the equation (16), let us separate the variables D and p by multiplying both sides by

$$\frac{dp}{p \cdot (A + B \cdot D)};$$

then, the equation takes the form

$$\frac{dD}{A + B \cdot D} = \frac{dp}{p}. \tag{17}$$

This equation is easy to integrate; the resulting solution is slightly different for $B = 0$ and $B \neq 0$.

6.1°. If $B = 0$, then integrating both parts of (17), we get $A^{-1} \cdot D = \ln(p) + C_1$ ($C_1, C_2, \ldots$ will denote constants). Hence,

$$D(p) = d'(p) = A \cdot \ln(p) + C_2,$$

and integrating again, we get

$$d(p) = A \cdot p \cdot \log(p) + C_2 \cdot p + C_3$$

for some constants C_i.

Thus, in this case, the function $d(p)$ is equivalent to the function $\pm p \cdot \ln(p)$ that corresponds to the Shannon index.

6.2°. If $B \neq 0$, then

$$\frac{dD}{A + B \cdot D} = \frac{d(D + c)}{A \cdot (D + c)},$$

where we denoted

$$c \stackrel{\text{def}}{=} \frac{A}{B},$$

hence

$$\frac{d(D + c)}{A \cdot (D + c)} = \frac{dp}{p}. \tag{18}$$

After integrating both parts of the equation (18), we get

$$A^{-1} \cdot \ln(D + c) = \ln(p) + C_1; \tag{19}$$

hence $\ln(D+c)=A\cdot\ln(p)+C_2$, and so, after exponentiating, we get $D+c=C_3\cdot p^A$. Thence,

$$d'(p)=D(p)=C_3\cdot p^A+C_4.$$

- If $A\neq -1$, we get

  $$d(p)=C_5\cdot p^{A+1}+C_4\cdot I+C_6.$$

  Thus, the function $d(p)$ is equivalent to the Rényi index.
- If $A=-1$, we similarly get

  $$d(p)=C_5\cdot\ln(p)+C_4\cdot I+C_6,$$

  in which case the function $d(p)$ is equivalent to $\pm\ln(p)$.

7°. In both cases $B=0$ and $B\neq 0$, the function $d(p)$ is equivalent to one of the functions listed in the formulation of the Proposition. theorem. The Proposition is thus proven.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

REFERENCES

[1] S. Braude and B. S. Low (eds.), *An Introduction to Methods and Models in Ecology, Evolution, and Conservation Biology*, Princeton University Press, Princeton, New Jersey, 2010.

[2] G. Cox, *General Ecology Laboratory Manual*, McGraw-Hill, New York, 2001.

[3] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[4] P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, MD, 1988.

[5] J. P. Gibbs and W. T. Martin, "Urbanization, technology, and the division of labor", *American Sociological Review*, 1962, Vol. 27, pp. 667–677.

[6] O. C. Herfindal, *Concentration in the U.S. Steel Industry*, PhD Dissertation, Columbia University, 1950.

[7] A. O. Hirshman, *National Power and the Structure of Foreign Trade*, Berkeley, 1945.

[8] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[9] O. Kosheleva, "Symmetry-group justification of maximum entropy method and generalized maximum entropy methods in image processing", In: G. J. Erickson, J. T. Rychert, and C. R. Smith (eds.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1998, pp. 101–113.

[10] D. R. Luce and H. Raiffa, *Games and Decisions, Introduction and critical survey*, John Wiley & Sons, Inc., New York, 1957.

[11] R. B. Myerson, *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA, 1991.

[12] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[13] A. Rényi, "On measures of entropy and information", *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, pp. 547–561.

[14] C. E. Shannon, "A mathematical theory of communication", *The Bell Systems Technical Journal*, 1948, Vol. 27, pp. 379–423 and 623–656.

[15] E. H. Simpson, "Measurement of diversity", *Nature*, 1949, Vol. 163, p. 688.

[16] P. Suppes, D. M. Krantz, R. D. Luce, and A. Tversky, *Foundations of Measurement. Vol. II. Geometrical, Threshold, and Probabilistic Representations*, Academic Press, San Diego, CA, 1989.

[17] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.