Journal of Physics Communications

CrossMark

Diversity in complex systems: measuring parts of the distribution to the whole

R Rajaram[1] ![ORCID] and B Castellani[2]

1 Department of Math. Sci., Kent State University, United States of America
2 Durham University, Durham, United Kingdom

E-mail: rrajaram@kent.edu

Abstract

Despite the utility of the Hill numbers, measuring diversity within complex systems and complex datasets, particularly regarding parts of a distribution to the whole—that is, different portions of diversity types to the entire system or dataset—is a challenging issue. In this paper, we attempt to relate the diversity of a part (or parts) of a distribution to the diversity of the whole. We derive this relationship for the Hill numbers ^qD, and use these results to further examine the effect of a freely varying type on the diversity of the whole distribution.

1. Measuring diversity

As identified by Sornette (Sornette 2009) and others (Newman 2005, Barabasi 2009), probability distributions are the 'first quantitative characteristics' of both complex systems and complex datasets (Sornette 2009, p. 2). In terms of diversity (both with regard to species richness and evenness), this makes them highly useful, as measurements on a wide range of complex systems and complex datasets are well approximated by their shape, particularly as the sample size and its diversity $n \rightarrow \infty$. For example, within the complexity sciences, the literature on measuring diversity is vast and includes a number of different mathematical formulations —Gini-Simpson index, Shannon entropy, Hill numbers— (Jost 2006, Leinster and Cobbold 2012, Chao and Jost 2015, Hsieh *et al* 2016, Pavoine and Marcon 2016, Jost 2018). The Hill numbers and Rényi entropies are also important measures used to derive quantum uncertainty relations in quantum theory and also the maximal entropy principle in statistical inference (see Jizba *et al* 2015, Jizba and Korbel 2018 and references therein).

Across these measures, the Hill numbers ^qD provide a highly useful way to measure the diversity of a distribution that computes the number of equivalent equiprobable types for a uniform distribution that maintains the same level of entropy (Rényi entropy to be precise) (macArthur 1965, Hill 1973, Peet 1974, Jost 2006, Gaggiotti *et al* 2018, Jost 2018). The parameter q favours the types with lower frequencies if $0 < q < 1$ and the types with higher frequencies for $q > 1$. For $q = 1$, ^1D weights each type proportional to their relative frequency of occurrence, and boils down to e^H where H is the Shannon entropy of the distribution.

Still, despite the utility of the Hill numbers, measuring diversity within complex systems and complex datasets, particularly regarding parts of a distribution to the whole—that is, different portions of diversity types to the entire system or dataset—is a challenging issue. In this paper, we attempt to relate the diversity of a part (or parts) of a distribution to the diversity of the whole. We derive this relationship for the Hill numbers ^qD, and use these results to further examine the effect of a freely varying type on the diversity of the whole distribution.

Our interest in this issue is based on a series of papers we have written on diversity in complex systems and complex datasets, in which we have introduced a modification of ^1D called *case-based entropy* C_c to more effectively measure the true diversity of some dataset of study, including the restriction of diversity within complex systems (Rajaram and Castellani 2012, 2014). As a measure, C_c is based on a modification of the Shannon-Wiener entropy H; and is inspired by Jost's paper on *Entropy and Diversity*, (Jost 2006), in which he states that: 'In physics, economics, information theory, and other sciences, the distinction between the entropy

Table 1. General dataset with complexity types x_i each having a probability p_i and a frequency f_i.

X	P	F
x_1	p_1	f_1
x_2	p_2	f_2
x_3	p_3	f_3
$\vdots$	$\vdots$	$\vdots$
x_J	p_J	f_J
$\vdots$	$\vdots$	$\vdots$
x_K	p_K	f_K

of a system and the effective number of elements of a system is fundamental. It is this latter number, not the entropy, that is at the core of the concept of diversity in biology' (p. 363).

To our knowledge such an approach, relative to the probability distributions of complex systems or complex datasets in general (particularly big data), has not been derived or explored before. As such, for us, Jost's quote—and, in turn, our development of Shannon-Wiener entropy H into C_c—provides one potentially useful way for advancing the statistical measurement of diversity within complex systems in general. In addition, we anticipate the use of our results in furthering our development of case-based entropy and deriving new properties of the same.

Remark 1.1. We note that the choice of q actually encodes the kind of correlation that exists between the parts of P_i of the distribution (Jizba and Korbel 2018). $q = 1$ means that the parts are independent and the *strong system independence condition* is satisfied i.e., disjoint parts are independent. Other choices of q will correspond to a sub or super exponential relationship between the number of distinguishable states and distinguishable sub-systems, as found in strongly correlated systems. So inherently, the choice of q actually chooses the kind of correlations that the partitions P_i have, and it is with that caveat that we endeavor to establish the relationship between diversity of the parts to the whole.

1.1. Purpose of current study

In terms of the current study, we consider a general probability distribution with a random variable X as shown in table 1 (signifying different types or categories), where x_i denotes the $i - th$ type, with probability p_i and frequency f_i. We ask the following question: if P_i denotes a partition of the indices $\{1, \ldots, K\}$ and $^qD_{P_i}$ represents the diversity of that partition, what is the relationship between $^qD_{P_i}$ and the diversity of all K types given by ^qD_K? And in particular, how does ^qD_K change if the probability of a single type p_K (say the K-th type) varies from 0 to 1. In other words, we are interested in the functional relationship of $^qD_K(p_K)$ as p_K varies from 0 to 1.

These results also address a key aspect of diversity and its relationship with the probability (or relative frequency) of the free type; namely, they show that as p_K increases, there exists an optimal level of frequency $\hat{p}_K$ so that the maximal diversity, which is equal to one more than the diversity of the known types i.e., $^qD_{K-1} + 1$, is attained, and the diversity starts to decrease for all probabilities $\hat{p}_K < p_K < 1$. Furthermore, there exists a frequency $\tilde{p}_K$ so that $^qD_K(\tilde{p}_K) = {}^qD_{K-1}$. In other words, when the probability of the K-th type equals $\tilde{p}_K$, the diversity of all K types is equal to the diversity of the $K - 1$ types. And, in some sense, at $p_K = \tilde{p}_K$, it is as though the K-th type does not appear to exist. Also, for all probabilities $\tilde{p}_K < p_K < 1$ the diversity starts decreasing and when $p_K = 1$, the K-th type dominates entirely leading to a total diversity of 1 i.e. $\lim_{p_K \to 1}^q D(p_K) = 1$.

We have organized the paper as follows. In section 2, we give a brief introduction to the idea of measurement of diversity. In section 3, we prove the two main theorems in the paper for ^1D_K and ^qD_K, $q \neq 1$ respectively. In section 4, we use the results from section 3 to investigate the function $^qD_K(p_K)$. And we finally conclude with a summary in section 5. At the outset, we mention that superscripts on the left are labels for diversities distinguishing the choices of q amongst Hill numbers. We attach an appendix to clarify some of the notation used, in addition to explaining it in the paper as well.

2. A formal introduction to diversity

Diversity, as a measure, counts the 'richness' (number of types) of a distribution in relation to the 'evenness' (equal probability of occurrence) of its diversity types (MacArthur 1965, Hill 1973, Peet 1974, Jost 2006, Rajaram and Castellani 2012, 2014). The intuition behind this definition is that if all of the types in the distribution occur

with the same probability, then diversity should simply be equal to the number of types K; and any deviation from uniformity in probabilities will always lead to a lower value of diversity.

Definition 2.1. Given an ordered set of types numbered as $i \in \mathbf{N}$ and their corresponding probabilities p_i, the diversity of the entire distribution ^qD_K for some complex system or dataset is defined as the number of equiprobable types needed to yield the same value of entropy ^qH.

Shannon entropy is defined as below:

$$^1H_K = -\sum_{l=1}^{K} p_l \ln(p_l) \tag{1}$$

Rényi entropy is defined as below:

$$^qH_K = \frac{1}{1-q} \ln\left(\sum_{l=1}^{K} p_l^q\right). \tag{2}$$

It was shown (MacArthur 1965, Hill 1973, Peet 1974, Jost 2006, Rajaram and Castellani 2012, 2014) that definition 2.1 implies that the total diversity ^qD_K is given by:

$$^1D_K = e^{^1H_K} = \Pi_{l=1}^{K} \frac{1}{p_l^{p_l}}, \tag{3}$$

and

$$^qD_K = e^{^qH_K} = \left(\sum_{i=1}^{K} p_i^q\right)^{\frac{1}{1-q}}, \quad q \neq 1. \tag{4}$$

Furthermore, we denote the diversity of any partition of indices of types in the following form $P_{\{1,i\}} = \{1,\dots,i\}$ as $^qD_{\{1,i\}}$. The partial diversity for such partition P_i can be written as follows:

$$^1D_{\{1,i\}} = \Pi_{l=1}^{i} \frac{1}{(p_{l\{1,i\}})^{p_{l\{1,i\}}}}; ^q D_{\{1,i\}} = \left(\sum_{l=1}^{i} (p_{l\{1,i\}})^q\right)^{\frac{1}{1-q}}, \quad q \neq 1, \tag{5}$$

where $p_{l\{1,i\}} = \frac{p_l}{\sum_{k=1}^{i} p_k} = \frac{f_l}{\sum_{k=1}^{i} f_k}$ are the marginal probabilities. We note that equations (3)–(5) can be rewritten in terms of the frequencies f_i as below:

$$^1D_K = \frac{\sum_{l=1}^{K} f_l}{\Pi_{l=1}^{K} f_l^{\left(\frac{f_l}{\sum_{l=1}^{K} f_l}\right)}}; ^1 D_{\{1,i\}} = \frac{\sum_{l=1}^{i} f_l}{\Pi_{l=1}^{i} f_l^{\left(\frac{f_l}{\sum_{l=1}^{i} f_l}\right)}}, \tag{6}$$

$$^qD_K = \frac{\left(\sum_{l=1}^{K} f_l^q\right)^{\frac{1}{1-q}}}{\left(\sum_{l=1}^{K} f_l\right)^{\frac{q}{1-q}}}; ^q D_{\{1,i\}} = \frac{\left(\sum_{l=1}^{i} f_l^q\right)^{\frac{1}{1-q}}}{\left(\sum_{l=1}^{i} f_l\right)^{\frac{q}{1-q}}}. \tag{7}$$

For an arbitrary disjoint partition P_i with indices not necessarily in order (we note that $P_{\{1,i\}}$ has indices in order from 1 to i for the types), the partial diversities in equation (5) can be rewritten as follows, with D_{P_i} denoting the diversity of the partition P_i:

$$^1D_{P_i} = \Pi_{l \in P_i} \frac{1}{(p_{lP_i})^{p_{lP_i}}}; ^q D_{P_i} = \left(\sum_{l \in P_i} (p_{lP_i})^q\right)^{\frac{1}{1-q}}, \quad q \neq 1, \tag{8}$$

where $p_{lP_i} = \frac{p_l}{\sum_{k \in P_i} p_k} = \frac{f_l}{\sum_{k \in P_i} f_k}$. Consequently the partial diversities in terms of frequencies in equations (6) and (7) can also be rewritten with arbitrary partitions with appropriate notation as follows:

$$^1D_{P_i} = \frac{\sum_{l \in P_i} f_l}{\Pi_{l \in P_i} f_l^{\left(\frac{f_l}{\sum_{l \in P_i} f_l}\right)}}, \tag{9}$$

$$ {}^qD_{P_i} = \frac{\left(\sum_{l \in P_i} f_l^q \right)^{\frac{1}{1-q}}}{\left(\sum_{l \in P_i} f_l \right)^{\frac{q}{1-q}}}. \tag{10}$$

3. Parts of the diversity distribution to the whole

Theorem 3.1. *Given a probability distribution similar to table 1, the diversity of the entire distribution ${}^qD_K$ for some complex system or dataset, and the diversities of disjoint parts ${}^qD_{P_i}$ and their respective cumulative probabilities c_{P_i} are related as follows:*

$$ {}^1D_K = \prod_{P_i \in \mathcal{P}} \left(\frac{{}^1D_{P_i}}{c_{P_i}} \right)^{c_{P_i}}, \tag{11}$$

and

$$ {}^qD_K = \left(\sum_{P_i \in P} c_{P_i}^q \, ({}^qD_{P_i})^{(1-q)} \right)^{\frac{1}{1-q}} \tag{12}$$

Proof 3.1.

- **The case q = 1**: Let us first look at the pooling of m disjoint distributions (i.e., none of the types repeat) with total frequencies equaling N_1, N_2 ,.., N_m respectively (i.e., $N = \sum_i f_i$ for each distribution) and diversities ${}^1D_1,{}^1 D_2,..,{}^1 D_m$ respectively. Let us consider a new distribution which pools these m distributions with total frequency $N_p = \sum_{l=1}^m N_l$. Let ${}^1D_p$ denote the diversity of the pooled distribution. Then, each of those m individual distributions are equivalent to a uniform distribution each with frequency $\frac{N_l}{{}^1D_l}$ for all the types within them. We have the following from equation (6):

$$ {}^1D_p = N_p e^{-\frac{\sum_{l=1}^m N_l \ln\left(\frac{N_l}{{}^1D_l}\right)}{N_p}} = \frac{N_p \prod_{l=1}^m ({}^1D_l)^{\frac{N_l}{N_p}}}{\prod_{l=1}^m N_l^{\frac{N_l}{N_p}}} = \frac{\prod_{l=1}^m ({}^1D_l)^{\frac{N_l}{N_p}}}{\prod_{l=1}^m \left(\frac{N_l}{N_p}\right)^{\frac{N_l}{N_p}}} = \prod_{l=1}^m \left(\frac{{}^1D_l}{p_l} \right)^{p_l}, $$

  where $p_l = \frac{N_l}{N_p}$ are the probabilities of occurence of each of the individual distributions assuming they are equally likely. Now, thinking of a given distribution as a pooling of its disjoint parts with indices in $P_i \in \mathcal{P}$—however, we may want to divide it—and changing the notation for the part P_i as the part l from above, we have that ${}^1D_l =^1 D_{P_i}$. The probability of the part of the distribution corresponding to indices in $P_i \in \mathcal{P}$ will then be $p_l = \sum_{r \in P_i} p_r = c_{P_i}$, the sum of all the probabilities of the individual types comprising the part P_i. Then the above equation's last step can be rewritten using the new notation to obtain equation (11).

- **The cases q $\neq$ 1**: The idea of the proof is the same i.e., start with pooling m disjoint distributions where the ${}^1D_l$'s are replaced now by ${}^qD_l$'s and so on. However the calculation from equation (7) for the pooling is different and as follows:

$$ {}^qD_p = \frac{\left(\sum_{l=1}^m N_l^q ({}^qD_l)^{1-q} \right)}{(N_p)^{\frac{q}{1-q}}} = \left(\sum_{l=1}^m \left(\frac{N_l}{N_p} \right)^q ({}^qD_l)^{(1-q)} \right)^{\frac{1}{1-q}} = \left(\sum_{l=1}^m (p_l)^q ({}^qD_l)^{(1-q)} \right)^{\frac{1}{1-q}}, $$

  and the result follows by the same renaming convention i.e. P_i as the part l as in the previous $q = 1$ case.

Remark 3.1. If we consider each part P_i in the derivation of the above theorem to be exactly one type i.e., $P_i = \{i\} \; \forall \, i = 1,...,K$, then ${}^qD_{P_i} = 1 \; \forall \, i = 1,...,K, \; q \in [0, \infty)$ and equation (11) reduces to equation (3), and equation (12) reduces to 4.

Remark 3.2. We can restrict ourselves to a portion of the distribution starting from $l = 1$ to say $l = k$. Then theorem 3.1 is true for the restriction for any sub-partition $\mathcal{P}_k$ of such a restriction. In other words, the result is true for a part of part of a distribution and for all nested sub-parts. In this case the probabilities will have to be re-normalised as $p_l = p_{l_{P_k}}$ and $c_l = c_{l_{P_k}}$, i.e., all probabilities and cumulative probabilities should be divided by the sum of all probabilities in the partition P_k. So this result is **self-similar** in nature.

Remark 3.3 (Example). Let us consider the following probability distribution in table 2: here $K = 5$. Let us consider the partition $P_1 = \{1, 2\}$, $P_2 = \{3, 4\}$ and $P_3 = \{5\}$. We can compute ${}^1D_{P_1} = 1.417,{}^1 D_{P_2} = 1.755,{}^1 D_{P_3} = 1$, and ${}^1D_5 = 3.747$. We can also compute the following: $c_{P_1} = \frac{9}{16}$, $c_{P_2} = \frac{1}{4}$ and $c_{P_3} = \frac{3}{16}$. The

Table 2. Example dataset for remark 3.3 with complexity types x_i each having a probability p_i and a frequency f_i.

X	P
x_1	$\frac{1}{16}$
x_2	$\frac{1}{2}$
x_3	$\frac{3}{16}$
x_4	$\frac{1}{16}$
x_5	$\frac{3}{16}$

following can be easily checked numerically:

$$^1D_5 = \left(\frac{^1D_{P_1}}{c_{P_1}}\right)^{c_{P_1}} \cdot \left(\frac{^1D_{P_2}}{c_{P_2}}\right)^{c_{P_2}} \cdot \left(\frac{^1D_{P_3}}{c_{P_3}}\right)^{c_{P_3}}.$$

Consequently if any three of $^1D_5, ^1D_{P_1}, ^1D_{P_2}, ^1D_{P_3}$ and the fourth is unknown (assuming $c_{P_1}, c_{P_2}, c_{P_3}, c_{P_4}$ are known), the fourth diversity (or partial diversity) can be computed.

Since entropy is the natural log of diversity, and diversity is the exponential of entropy, any result related to diversity is also a result related to entropy and vice-versa. We finish this section by stating a corollary of theorem 3.1 which relates the entropy of parts of a distribution to the total entropy.

Corollary 3.1. *Given a probability distribution similar to table 1, the entropy of the entire distribution ^qH_K for some complex system or dataset, and the entropies of disjoint parts $^qH_{P_i}$ and their respective cumulative probabilities c_{P_i} are related as follows:*

$$^1H_K = \sum_{P_i \in \mathcal{P}} (c_{P_i}(\,^PH_{P_i} - \ln(c_{P_i}))), \tag{13}$$

and

$$^qH_K = \frac{1}{1-q} \ln\left(\left(\sum_{P_i \in P} c_{P_i}^q (e^{\,^qH_{P_i}\cdot(1-q)})\right)\right). \tag{14}$$

Proof 3.2. The proof follows by taking natural logarithm of equations (11) and (12) respectively.

4. Diversity and the probability of the free type

As the reader may recall, our results also address a key aspect of diversity and its relationship with the probability (or relative frequency) of the free type. To demonstrate, we apply theorem 3.1 to investigate the effect of changing the probability of the last type K on the total diversity of ^qD_K. Changing the probability p_K will mean re-normalizing the probabilities of the remaining $K-1$ types to make the total probability $= 1$. We note that, equivalently, we could simply rewrite all formulas in terms of actual frequencies and change the frequency f_K of the K-th type. Using probabilities, however, will allow us to draw graphs between 0 and 1 which is a bit more convenient (and also visually useful for investigation) to compare the variations for different choices of q.

We state a corollary to theorem 3.1 which can be easily proved by choosing $P_1 = \{1,\ldots,(K-1)\}$ and $P_2 = \{K\}$.

Corollary 4.1. *Given a probability distribution like in table 1, the diversity of the first $(K-1)$ types $^qD_{K-1}$, the diversity of the whole distribution ^qD_K and the probability of the K-th type p_K are related as follows:*

$$^1D_K = \left(\frac{^1D_{K-1}}{(1-p_K)}\right)^{(1-p_K)} \left(\frac{1}{p_K}\right)^{p_K}, \tag{15}$$

and

$$^qD_K = (^qD_{K-1}^{(1-q)}(1-p_K)^q + p_K^q)^{\frac{1}{1-q}}. \tag{16}$$

Proof 4.1. The proof follows by applying theorem 3.1 to the partition $P_1 = \{1,\ldots,(K-1)\}$ and $P_2 = \{K\}$.
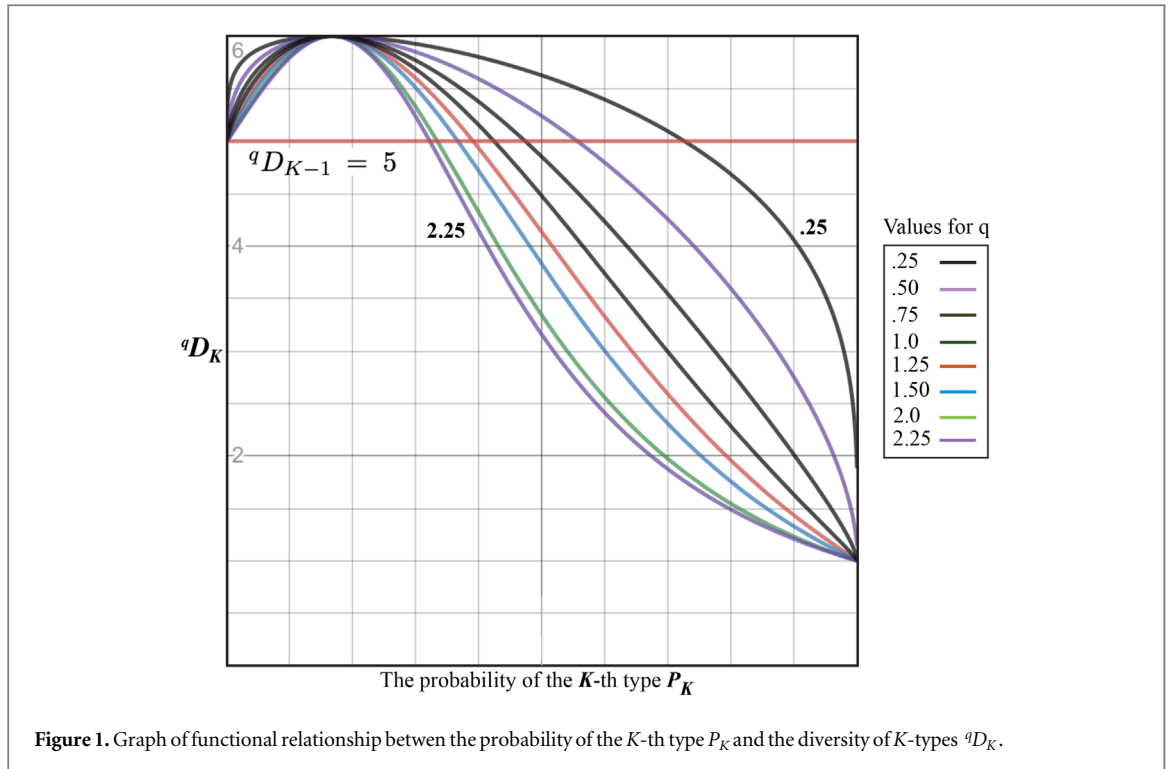
Figure 1. Graph of functional relationship betwen the probability of the K-th type P_K and the diversity of K-types ^qD_K.

Remark 4.1. It is to be noted that even though equations (15) and (16) show the explicit relationship between ^qD_K and p_K for all $q \in [0, 1]$, $^qD_{K-1}$ is still completely independent of p_K. This is because while computing marginal probabilities $p_{l\{1,(K-1)\}}$ (even if we choose to use the already computed probabilities $p_{l\{1,K\}}$!) the dependence on p_K cancels out.

$$p_{l\{1,(K-1)\}} = \frac{p_{l\{1,K\}}}{\sum_{k=1}^{K-1} p_{k\{1,K\}}} = \frac{\left(\frac{p_l}{\sum_{k=1}^{K} p_k}\right)}{\sum_{l=1}^{K-1} \frac{p_l}{\sum_{k=1}^{K} p_k}} = \frac{p_l}{\sum_{k=1}^{K-1} p_k}.$$

Equations (15) and (16) give us a direct relationship between the probability of the K-th type and the total diversity. For the specific choice of $K = 6$ and $^qD_5 = 5$, we plot the function $^qD_K(p_K)$ for various choices of q. We note that the choice of K and $^qD_{K-1}$ has no bearing on the shape of the functional relationships that we are discussing in this section.

We make the following observations from the graph 'in figure 1':

(1) **Maximum**: Diversity is maximized for all choices of q when $\hat{p}_K = \frac{1}{(^qD_{K-1}+1)}$ and the maximum value of diversity is equal to $^qD_{K-1} + 1$ i.e., the K-th type is counted as 1 at maximum. We note that this can be proved separately by Calculus based techniques. (See theorem A.1 in the appendix).

(2) **Start**: When $p_K = 0$, as expected, the diversity satisfieds $^qD_K(0) = {}^qD_{K-1}$ as expected i.e., in the absence of the K-th type, the diversity is simply equal to the diversity of the remaining $K - 1$ types. This is also equivalent to the Shannon-Kinchin axiom 2 (Jizba and Korbel 2019).

(3) **Point of diminishing return**: We note that there exists a probability $\frac{1}{(^qD_{K-1}+1)} < \tilde{p}_K < 1$ where $^qD_K(\tilde{p}_K) = {}^qD_{K-1}$ i.e., there exists a probability of the K-th type where the total diversity of K-types is equal to the diversity of $K - 1$ types. So in a sense, at $p_K = \tilde{p}_K$, it is as though the K-th type does not add to the total diversity at all. For all probabilities $\tilde{p}_K < p_K < 1$, the diversity starts to decrease to a value less than the diversity of the $K - 1$ types. Hence, for $p_K > \tilde{p}_K$ the addition of the K-th type actually starts to decrease the total diversity, which justifies our terminology that $p_K = \tilde{p}_K$ is the point of diminishing returns. Another way to state this is to say that the K-th type starts to dominate the total diversity for $p_K > \tilde{p}_K$. There is no explicit analytical form for $\tilde{p}_K$, as it is the solution of an implicit nonlinear equation. However, its existence can be proved using Calculus based techniques. (See corollary A.1 in the Appendix)

(4) **Variation of point of diminishing return:** As q increases, the point of diminishing return $\tilde{p}_K$ gets monotonically smaller.

(5) **Asymptotic behavior at as $\mathbf{p}_K \to 1$:** As $p_K \to 1$, the K-th type completely dominates and the diversity satisfies $^qD_K \to 1$ i.e., it is as though the previous $K-1$ types do not even exist! This is also equivalent to the Shannon-Kinchin axiom 2 (Jizba and Korbel 2019).

5. Conclusion

In this paper, relative to the distribution of diversity in complex systems and datasets, we have derived an explicit statistical relationship between the diversity of part of a distribution and the whole. Relative to our findings, the following are our conclusions:

(1) **Main result:** The relationship in theorem 3.1 in equations (11) and (12) provide an explicit functional relationship between the diversity of a partition P_i of a distribution (given by $^qD_{P_i}$ and the total diversity ^qD_K of the K types. To our knowledge this is the first time such a relationship has been derived using this approach.

(2) At the bottom of the hierarchy, if we consider each type as a part itself, then we can recover the original formulas given by 3 and 4.

(3) **Self-similar nature:** The relationship between the diversity of parts to the whole is self-similar in nature i.e., the same relationship exists between a part of the distribution and its own parts as well, as explained in remark 3.2.

(4) **Entropy of parts:** The relationship between the diversity of parts to the whole leads to a similar fractal relationship between entropy of parts and the whole as seen in corollary 3.1.

(5) **Usefulness:** The main result in the paper (Theorem 11) can be used to explore the relationship between the probability of occurrence of a part of a distribution and the diversity of the whole distribution for a variety of complex system and datasets. As an example, we have shown how it can be used to explore the dependence of the total diversity ^qD_K of K types on the probability of the K-th type and found some interesting revelations along the way such as the point of diminishing returns. Complexity is all about relationships (or lack thereof) between parts and whole. Theorem 3.1, corollary 3.1 and corollary 4.1 explicitly relate the diversity (or entropy) of parts of a distribution to the diversity of the entire distribution. This is important because it allows us to directly compute the change in the diversity of the entire distribution if the diversity of the parts change due to complex internal mechanisms.

(6) **Future work:** In our future work we will try to use the main result in this paper to explore the relationship between case-based entropy of a given distribution and its shape. We believe this will be a useful step towards using case-based entropy as a tool to further explore the regions of the distribution that contribute to the inequality of diversity. For example, in a variety of probability distributions it could prove useful to pinpoint and quantify the contribution of parts of a distribution to the overall diversity.

Appendix A. Proof of graph properties

Theorem A.1. $^qD_K(p_K)$ *attains a maximum at* $\hat{p}_K = \frac{1}{^qD_{K-1}+1}$ *and the maximum diversity is given by* $^qD_K(\hat{p}_K) = {}^q D_{K-1} + 1$.

Proof A.1. We show that $^qD_K'(\hat{p}_K) = 0$ and $^qD_K''(\hat{p}_K) < 0$ for all q. We show the calculation for $q \neq 1$ first.

$$^qD_K = ({}^qD_{K-1}^{(1-q)}(1-p_K)^q + p_K^q)^{\frac{1}{1-q}}$$

$$\Rightarrow {}^qD_k'(p_K) = \frac{1}{1-q}({}^qD_{K-1}^{1-q}(1-p_K)^q + p_K^q)^{\frac{q}{1-q}} \cdot ({}^qD_{K-1}^{1-q}(1-p_K)^q + p_K^q)'$$

$$^qD_k'(p_K) = \frac{q}{1-q}({}^qD_{K-1}^{1-q}(1-p_K)^q + p_K^q)^{\frac{q}{1-q}} \cdot [-{}^qD_{K-1}^{1-q} \cdot (1-p_K)^{q-1} + p_K^{q-1}.]$$

$$^qD_k'(p_K) = 0 \Rightarrow -{}^qD_{K-1}^{1-q} \cdot (1-p_K)^{q-1} + p_K^{q-1} = 0 \Rightarrow \hat{p}_K = \frac{1}{^qD_{K-1}+1}, {}^q D_k(\hat{p}_K) = {}^q D_{K-1} + 1.$$

Also, $\hat{p}_K$ is the only solution for ${}^qD'_K(\hat{p}_K) = 0 \quad \forall\, p_K \in [0, 1]$.

$$
{}^qD''_K(p_K) = \frac{q}{1-q}({}^qD_{K-1}^{1-q}(1-p_K)^q + p_K^q)^{\frac{q}{1-q}}.
$$

$$
\left[\frac{q^2}{1-q} \frac{(-{}^qD_{K-1}^{1-q}\cdot(1-p_K)^{q-1} + p_K^{q-1})^2}{{}^qD_{K-1}^{1-q}(1-p_K)^q + p_K^q} + (q-1)\cdot({}^qD_{K-1}^{1-q}\cdot(1-p_K)^{q-2} + p_K^{q-2}.) \right]
$$

$$
{}^qD''_K(\hat{p}_K) = -q({}^qD_{K-1}^{1-q}(1-\hat{p}_K)^q + \hat{p}_K^q)^{\frac{q}{1-q}}\cdot({}^qD_{K-1}^{1-q}\cdot(1-\hat{p}_K)^{q-2} + \hat{p}_K^{q-2}) < 0
$$

Next we show the same calculations for $q = 1$:

$$
{}^1D_K = \left(\frac{{}^1D_{K-1}}{(1-p_K)} \right)^{(1-p_K)} \left(\frac{1}{p_K} \right)^{p_K}
$$

$$
\Rightarrow \ln({}^1D_K(p_K)) = (1-p_K)\ln\left(\frac{{}^1D_{K-1}}{1-p_K} \right) - p_K\ln(p_K)
$$

$$
\Rightarrow {}^1D'_K(p_K) = {}^1D_K(p_K)\left[\frac{{}^1D_{K-1}}{1-p_K} - \ln(p_K). \right]
$$

$$
{}^1D'_K(p_K) = 0 \Rightarrow \frac{{}^1D_{K-1}}{1-p_K} - \ln(p_K) = 0 \Rightarrow \Rightarrow \hat{p}_K = \frac{1}{{}^1D_{K-1}+1}, {}^1D_k(\hat{p}_K) = {}^1D_{K-1} + 1.
$$

$$
{}^1D''_K(p_K) = -{}^1D'_K(p_K)\ln\left(\frac{p_K^1 D_{K-1}}{1-p_K} \right) - {}^1D_K(p_K)\left(\frac{1-p_K}{p_K^1 D_{K-1}} \right)\left(\frac{p_K^1 D_{K-1}}{1-p_K} \right)'
$$

$$
= -{}^1D'_K(p_K)\ln\left(\frac{p_K^1 D_{K-1}}{1-p_K} \right) - \frac{{}^1D_K(p_K)}{p_K(1-p_K)}
$$

$$
\Rightarrow {}^1D''_K(\hat{p}_K) = -{}^1D'_K(\hat{p}_K)\ln\left(\frac{\hat{p}_K^1 D_{K-1}}{1-\hat{p}_K} \right) - \frac{{}^1D_K(\hat{p}_K)}{\hat{p}_K(1-\hat{p}_K)} = -\frac{{}^1D_K(\hat{p}_K)}{\hat{p}_K(1-\hat{p}_K)} < 0.
$$

In addition, $\hat{p}_K$ is the only critical point of ${}^1D_K(p_K)\forall\, p_K \in (0, 1)$. This in addition to the fact that ${}^qD_K(0) = {}^qD_{K-1}$ and ${}^qD_K(1) = 1 \quad \forall\, q \in (0, \infty)$ means that $p_K = \hat{p}_K$ is the absolute maximum of the function ${}^qD_K(p_K) \quad \forall\, p_K \in [0, 1]$ and $\forall\, q \in (0, \infty)$. This proves the theorem.

Corollary A.1. *There exists a* $\tilde{p}_K \in (\hat{p}_K, 1)$ *so that* ${}^qD_K(\tilde{p}_K) = {}^qD_{K-1} \quad \forall\, q \in (0, \infty)$ *and* ${}^qD_K(p_K) < {}^qD_{K-1}$ $\forall\, p_K \in (\tilde{p}_K, 1)$.

Proof A.2. The proof follows from the fact that ${}^qD_K(p_K)$ attains an absolute maximum at $p_K = \hat{p}_K$ and is decreasing on $p_K \in (\hat{p}_K, 1)$ and ${}^qD_K(1) = 1$. Hence, by continuity of ${}^qD_K(p_K)$ and the intermediate value theorem the graph has to pass through a point $p_K = \tilde{p}_K$ so that ${}^qD_K(\tilde{p}_K) = {}^qD_{K-1}$ before monotonically decreasing to ${}^qD_K(1) = 1$.

Appendix B. Notation

(1) K: The number of types in a distribution.

(2) D_K: Diversity of the entire distribution i.e., all K types.

(3) $\mathcal{P}$: An ascending disjoint partition of the set of indices $\{1,.., K\}$ such that every element $P_i \in \mathcal{P}$ satisfies the property that $i < j \Rightarrow \max P_i < \max P_j$. In other words, the partition preserves the ordering of the numbers $\{1,.., K\}$. In particular, the member $\{i, (i+1),.., j\}$ denotes the types in the distribution between indices i and j and will be denoted by $\{i, j\}$.

(4) D_{P_i}: Diversity of the part of the distribution corresponding to indices in $P_i \in \mathcal{P}$.

(5) $c_{P_i} = \sum_{l\in P_i} p_l$: sum of probabilities of types in the part of the distribution corresponding to indices in $P_i \in \mathcal{P}$.

(6) $p_{l\{1,i\}} = \frac{p_l}{c_{\{1,i\}}} = \frac{p_l}{\sum_{k=1}^{i} p_k} = \frac{f_l}{\sum_{k=1}^{i} f_k}$: marginal probabilities for the first i types $l = 1,.., i$. Same definition for p_{lP_i} for any partition $P_i \in \mathcal{P}$ i.e., $p_{lP_i} = \frac{p_l}{c_{P_i}} = \frac{p_l}{\sum_{k\in P_i} p_k} = \frac{f_l}{\sum_{k\in P_i} f_k}$.

(7) $c_{l\{1,i\}} = \frac{c_l}{c_{\{1,i\}}} = \frac{c_l}{\sum_{k=1}^{i} p_k}$: marginal cumulative probabilities for the first i types $l = 1,.., i$. Same definition for c_{lP_i} for any partition $P_i \in \mathcal{P}$ i.e., $c_{lP_i} = \frac{c_l}{c_{P_i}} = \frac{c_l}{\sum_{k \in P_i} p_k}$. So in general, whenever there is a partition P_i as a subscript, it means that we are dividing the probability (or cumulative probability) in the base by c_{P_i}.

ORCID iDs

R Rajaram ● https://orcid.org/0000-0003-1680-6706

References

Barabasi A L 2009 Scale-free networks: a decade and beyond *Science* **325** 412

Chao A and Jost L 2015 Estimating diversity and entropy profiles via discovery rates of new species *Methods in Ecology and Evolution* **6** 873–82

Gaggiotti O E, Chao A, Peres-Neto P, Chiu C-H, Edwards C, Fortin M-J, Jost L, Richards C M and Selkoe K A 2018 Diversity from genes to ecosystems: a unifying framework to study variation across biological metrics and scales *Evolutionary Applications* **11** 1176–1193

Hill M 1973 Diversity and evenness: a unifying notation and its consequences *Ecology* **54** 427–32

Hsieh T C, Ma K H and Chao A 2016 Inext: an r package for rarefaction and extrapolation of species diversity (h ill numbers) *Methods in Ecology and Evolution* **7** 1451–6

Jizba P, Dunningham J A and Joo J 2015 Role of information theoretic uncertainty relations in quantum theory *Ann. Phys.* **355** 87–114

Jizba P and Korbel J 2018 Maximum entropy principle in statistical inference: a case for non-Shannonian entropies *Phys. Rev. Lett.* **122** 120601 (1-6)

Jizba P and Korbel J 2019 When Shannon and Khinchin meet Shore and Johnson: equivalence of information theory and statistical inference axiomatics *Physical Review* E https://arxiv.org/abs/1909.12677

Jost L 2006 Entropy and diversity *Oikos* **113** 363–75

Jost L 2018 What do we mean by diversity? The path towards quantification *Mètode Science Studies Journal-Annual Review* **9** 55–61

Leinster T and Cobbold C A 2012 Measuring diversity: the importance of species similarity *Ecology* **93** 477–89

MacArthur R 1965 Patterns of species diversity *Biological Review* **40** 510–33

Newman M 2005 Power laws, pareto distributions and zipf's law *Contemp. Phys.* **46** 323–51

Pavoine S, Marcon E and Ricotta C 2016 'Equivalent numbers' for species, phylogenetic or functional diversity in a nested hierarchy of multiple scales *Methods in Ecology and Evolution* **7** 1152–63

Peet R 1974 The measurement of species diversity *Annual Review of Ecological Systems* **5** 285–307

Rajaram R and Castellani B 2012 Modeling complex systems macroscopically: case/agent-based modeling, synergetics and the continuity equation *Complexity* **18** 8–17

Rajaram R and Castellani B 2014 The utility of non-equilibrium statistical mechanics, specifically transport theory, for modeling cohort data *Complexity* **20** 45–57

Sornette D 2009 Probability distributions in complex systems *Encyclopedia of Complexity and Systems Science* ed R A Meyers (Berlin: Springer) pp 7009–24