Journal of Physics Communications

PAPER • OPEN ACCESS

On the mathematical quantification of inequality in probability distributions

To cite this article: R Rajaram *et al* 2024 *J. Phys. Commun.* **8** 085002

View the article online for updates and enhancements.

Journal of Physics Communications

CrossMark

OPEN ACCESS

RECEIVED
6 June 2024

REVISED
26 July 2024

ACCEPTED FOR PUBLICATION
1 August 2024

PUBLISHED
14 August 2024

PAPER

On the mathematical quantification of inequality in probability distributions

R Rajaram[1] , **N Ritchey**[1] and **B Castellani**[2]

[1] Department of Math. Sci., Kent State University, United States of America
[2] Durham Research Methods Centre, Durham University, Durham, United Kingdom

E-mail: rrajaram@kent.edu

Keywords: power law, case-based entropy, Shannon entropy

Abstract

A fundamental challenge in the study of probability distributions is the quantification of inequality that is inherently present in them. Some parts of the distribution are more probable and some others are not, and we are interested in the quantification of this inequality through the lens of mathematical diversity, which is a new approach to studying inequality. We offer a theoretical advance, based on case-based entropy and slope of diversity, which addresses inequality for arbitrary probability distributions through the concept of mathematical diversity. Our approach is useful in three important ways: (1) it offers a universal way to measure inequality in arbitrary probability distributions based purely on the entropic uncertainty that is inherent in them and nothing else; (2) it allows us to compare the degree of inequality of arbitrary parts of any distribution (not just tails) and entire distributions alike; and (3) it can glean out empirical rules similar to the 80/20 rule, not just for the power law but for any given distribution or its parts thereof. The techniques shown in this paper demonstrate a more general machinery to quantify inequality, compare the degree of inequality of parts or whole of general distributions, and prove or glean out empirical rules for general distributions based on mathematical diversity. We demonstrate the utility of this new machinery by applying it to the power law, the exponential and the geometric distributions. The $60-40$ rule of restricted diversity states that 60 percent or more of cases following a power law (or more generally a right skewed distribution) reside within 40 percent or less of the lower bound of Shannon equivalent equi-probable (SEE) types as measured by case-based entropy. In this paper, we prove the $60-40$ rule for power law distributions analytically. We also show that in all power law distributions, the second half of the distribution is at least 4 times more uniformly distributed as the first. Lastly, we also show a scale-free way of comparing probability distributions based on the idea of mathematical diversity of parts of a distribution. We use this comparison technique to compare the exponential and power law distribution, and obtain the exponential distribution as an entropic limit of the power law distribution. We also demonstrate that the machinery is applicable to discrete distributions by proving a general result regarding the comparison of parts of the geometric distribution.

1. Introduction

A very prominent example of a distribution that follows the power law is the distribution of wealth, where it is well known that very few people have a lot of wealth. This inequality of distribution is captured in the so called Pareto principle, which is an empirical rule that states that more than 80 percent of the total wealth is situated in the richest 20 percent of people. In (Newman 2005), a rich exposition of the power law can be found, in addition to an analytical proof of the Pareto principle, and also techniques to empirically detect power laws from data. The Pareto principle emphasizes the severity of inequality in distribution of wealth, and is a fundamental principle in its own right. However, this got us thinking along the following lines:

(1) Is there a universal way to measure inequality based purely on the entropic uncertainty that is inherent to probability distributions and nothing else?

(2) Can we make it so this new method allows us to compare parts of a distribution and entire distributions alike?

(3) Can we use this technique to glean out empirical rules similar to the Pareto principle, not just for the power law but for any given distribution or its parts thereof?

We strongly believe that the answer to all three of the above questions is yes and is provided by our recent discovery of case-based entropy and slope of diversity curves (Rajaram *et al* 2023), where a novel method of comparing the diversity of a part P (or whole) of a distribution was constructed based on the ratio $\frac{D_P}{c_P}$. Here, D_P is the mathematical (or entropic) diversity (or the number of Shannon equivalent equi-probable or SEE parts) and c_P is the cumulative probability of the part P, both of which we will explain shortly. We contend, that comparing the degree of uniformity (which we will redefine as degree of inequality later in section 3) of parts (or whole) of a distribution provides a scale free way to compare probability distributions or their parts in terms of how equally distributed the random variable is, on those parts. Given a part P (or the whole) of a continuous distribution, the number D_P gives us the support of an equivalent uniform distribution that will have the same conditional entropy as the original part P i.e., if we were to replace the part P with a uniform distribution without losing any of the entropic uncertainty of the part P, then that abstract equivalent distribution will have a support of D_P for the portion of the random variable X_P over the part P.

In this paper, we will apply our newly formulated theory of mathematical diversity of parts of a distribution (Rajaram *et al* 2023, 2024) to the power law distribution to demonstrate that the case-based entropy curve and the slope of diversity curves for the power law reveal much more about the degree of uniformity of parts of the power law. In particular, we will analytically prove the 60–40 rule that was empirically observed in right-tailed distributions in (Castellani and Rajaram 2016), for the power law. In addition, we will show that the second half of the power law distribution which contains the majority of the tail, is at least 4 times as uniformly distributed as the first half—a result that has not been shown before. We also provide a new link between the power law and exponential distribution through the lens of the slope of diversity curve. Specifically, we show that slope of diversity curve of the exponential distribution can be obtained as an entropic limit of the corresponding curves for the power law distribution as the parameter α goes to ∞. Finally, we also demonstrate the use of our machinery to discrete distributions by proving a general result relating the degree of inequality of parts of the geometric distribution .

The paper is organized as follows: In section 2, we will introduce the main ideas behind mathematical diversity. In section 3, we explain how the ratio $\frac{D_P}{c_P}$, in addition to being a quantification of the degree of uniformity of the part P, is also a quantification of the degree of inequality of a given part P in a distribution. In section 4, we recall some results related to the power law that were proved in (Rajaram *et al* 2024). These results will be used later on to derive properties of the degree of uniformity (or inequality) of parts of the power law distribution. In section 4.1, we derive an explicit analytical expression for the case-based entropy curve for the power law distribution and analytically prove the $60-40$ rule for power law distributions that was empirically observed in (Castellani and Rajaram 2016). In section 5 we prove some interesting results that compare the degree of uniformity of the first half of the power law distribution to the second half. In section 6 we do a novel comparison of the power law and the exponential distributions and show that the distribution of degree of uniformity of the exponential distribution as measured by its slope of diversity curve can be obtained as an entropic limit of the corresponding slope of diversity curve of the power law distribution as the parameter α tends to ∞. In section 7, we demonstrate the applicability of our machinery to discrete distributions by proving a general result on the comparison of degree of uniformity of parts for the geometric distribution. We finish the paper with some conclusions in section 8.

2. A formal introduction to diversity: background material

Diversity is frequently used as a way to assess the richness or number of categories in a distribution, as well as its evenness which denotes the equal probability of each type of diversity appearing, as underscored in numerous studies (Jost 2006, MacArthur 1965, Hill 1973, Peet 1974). For more on the exposition of diversity and distributions we refer the reader to (Chao and Jost 2015, Hsieh *et al* 2016, Jost 2006, 2018, Leinster and Cobbold 2012, Pavoine and Marcon 2016). The idea of diversity is based on the premise that if all the K categories in a discrete probability distribution have an equal chance of happening, then the diversity should be equal to the number of categories K. Similarly, for a continuous distribution, the diversity of a uniform distribution is simply the Lebesgue measure of its support. Conversely, any deviation from uniformity in probabilities will inevitably result in a reduction in diversity. We recall a few definitions and theorems that are valid for mathematical diversity of continuous distributions first. Similar equivalent definitions are true for discrete distributions as well as seen in (Rajaram *et al* 2023, 2024).

Definition 2.1. (Shannon Diversity corresponding to $q = 1$ for Hill numbers) Given a continuous random variable X with support (a,b) (with $a = -\infty$ and $b = +\infty$ allowed) and its probability density $p(x)$, the diversity of the entire distribution ${}^{1}D_{(a,b)}$ is defined as the length of the support of an equivalent uniform distribution that yields the same value of Shannon entropy H.

Shannon entropy is defined as below:

$$H_{(a,b)} = -\int_{(a,b)} p(x)\ln(p(x))dx. \tag{1}$$

It was shown (Jost 2006, MacArthur 1965, Hill 1973, Peet 1974) that definition 2.1 implies that the total diversity ${}^{1}D_{(a,b)}$ is given by:

$${}^{1}D_{(a,b)} = e^{H_{(a,b)}}. \tag{2}$$

We will only consider the case $q = 1$ for the Hill numbers and hence, we will omit the left superscript of 1 while referring to the diversity as D. The main advantage of using Shannon entropy corresponding to $q = 1$, as stated in (Rajaram *et al* 2023, 2024), is because both the richness and evenness of probability distributions are equally weighted with this choice of the Hill number. No other choice of Hill numbers or other measures of diversity satisfy this unique balancing property. Furthermore, Shannon entropy has the intuition uncertainty in probability distributions, and as we will see later, mathematical diversity is an equivalent equi-probable reformulation of a part or whole of a distribution that maintains the entropic uncertainty of said part or whole. We re-state the diversity of parts theorem for continuous distributions below.

Theorem 2.1. *Let* p(x) *be a probability density function (pdf) on* (a,b)*, with $a = -\infty$ and $b = +\infty$ permitted. Let $P = \bigcup_i P_i$ be a disjoint partition of a part $P \subset (a, b)$. Then the following is true:*

$$\left(\frac{D_P}{c_P}\right)^{c_P} = \prod_{P_i \in P} \left(\frac{D_{P_i}}{c_{P_i}}\right)^{c_{P_i}}. \tag{3}$$

We make some definitions to establish some notation to state our next theorem.

Definition 2.2. We define

$$A_P = \frac{D_P}{c_P \cdot D_{(a,b)}} \quad \text{and} \quad A_{P_i} = \frac{D_{P_i}}{c_{P_i} \cdot D_{(a,b)}} \tag{4}$$

to be the average case-based entropy per unit cumulative frequency for the part P and the sub-part P_i respectively.

Definition 2.3. Let $P = (a, x)$ be a part for a continuous probability distribution on (a,b), with $a = -\infty$ and $b = +\infty$ allowed. The graph of $c_{(a,x)}$ on the x-axis versus $c_{(a,x)} * \ln(A_{(a,x)})$ on the y-axis is defined as the slope of diversity curve. Also, the slope of the secant joining the points $(c_{(a,x_1)}, c_{(a,x_1)} * \ln(A_{(a,x_1)}))$ and $(c_{(a,x_2)}, c_{(a,x_2)} * \ln(A_{(a,x_2)}))$ on the slope of diversity curve is denoted by $S_{(x_1,x_2)}$.

We next define the degree of uniformity of a part $P = (x_1, x_2)$, which we will later redefine as the degree of inequality of the part P in section 3.

Definition 2.4. Let $P = (x_1, x_2)$ be a part for a continuous probability distribution on (a,b), with $a = -\infty$ and $b = +\infty$ allowed. The ratio $\frac{D_{(x_1,x_2)}}{c_{(x_1,x_2)}}$ is termed as degree of uniformity of the part $P = (x_1, x_2)$.

In (Rajaram *et al* 2023), the validity of the ratio $\frac{D_P}{c_P}$ as a quantitative measure of the degree of uniformity of a part P of a discrete distribution was established. The slope of diversity curve was shown to be useful to compute and compare the degrees of uniformity of parts of a distribution in (Rajaram *et al* 2023) by comparing the slopes of secants of the corresponding parts from this curve. We recall the version of that theorem for continuous distributions below.

Theorem 2.2. *Let $p(x)$ be a probability density function (pdf) on* (a,b)*, with $a = -\infty$ and $b = +\infty$ permitted. Let (x_1, x_2) and (x_3, x_4) be parts that are subsets of (a,b). Then the following are true:*

$$\frac{D_{(x_1,x_2)}}{c_{(x_1,x_2)}}\begin{pmatrix}<\\=\\>\end{pmatrix}\frac{D_{(x_3,x_4)}}{c_{(x_3,x_4)}}\Longleftrightarrow S_{(x_1,x_2)}\begin{pmatrix}<\\=\\>\end{pmatrix}S_{(x_3,x_4)}. \tag{5}$$

$$\frac{D_{(x_1,x_2)}}{c_{(x_1,x_2)}} = De^{S_{(x_1,x_2)}}. \tag{6}$$

Remark 2.1. Theorem 2.1 relates the degree of uniformity $\frac{D_P}{c_P}$ of a given part P of a continuous distribution as the weighted geometric mean of the degree of diversity of $\frac{D_{P_i}}{c_{P_i}}$ of its sub-parts P_i with the cumulative probabilities c_{P_i} as the weights. Theorem 2.2 means that when comparing the slopes of secants $S_{(x_1,x_2)}$ of the slope of diversity curve, we are also comparing the degrees of uniformity in the parts (x_1, x_2) and (x_3, x_4). It also means that we can compute the degree of uniformity $\frac{D_{(x_1,x_2)}}{c_{(x_1,x_2)}}$ of an arbitrary part $P = (x_1, x_2)$ directly from the slope of secant $S_{(x_1,x_2)}$ of the slope of diversity curve. This is the main importance of the two results in this section.

Definition 2.5. Let $p(x)$ be a probability density function (pdf) on (a,b), with $a = -\infty$ and $b = +\infty$ permitted. The graph of $p(x)$ is denoted by g_1. The graph of $c_{(a,x)}$ on the x-axis versus $C_{(a,x)} = \frac{D_{(a,x)}}{D_{(a,b)}}$ on the y-axis is called the case-based entropy curve. We denote the case-based entropy curve by g_2. The graph of $c_{(a,x)}$ versus $c_{(a,x)} \cdot \ln(A_{(a,x)})$ also known as the slope of diversity curve is denoted by g_3.

We finish this section by simply stating that several important results relating the equivalence of the curves g_1, g_2 and g_3 and the method of reconstructing the original probability distribution from them have been proved in (Rajaram *et al* 2023, 2024).

3. The ratio $\frac{D_P}{c_P}$ as a quantification of degree of inequality of a given part P

In this section, we first recall the thought process behind the concept of *degree of uniformity* and then we establish the equivalence of this with *degree of inequality*.

3.1. $\frac{D_P}{c_P}$ as degree of uniformity

In (Rajaram *et al* 2023, 2024), we introduced the ratio $\frac{D_P}{c_P}$ as a way to measure the degree of uniformity of a part P of a given probability distribution irrespective of whether it is continuous or discrete. The main idea was that the part P of the given distribution can be redrawn as a Shannon Equivalent Equi-probable uniform distribution with a support of length D_P and a uniform probability of $\frac{c_P}{D_P}$. This equivalence allowed us to interpret the ratio $\frac{D_P}{c_P}$ as the SEE extent per unit cumulative frequency coming from the part P. Hence, the larger this ratio is, the larger the uniformity of the part P is.

We computed the ratio $\frac{D_P}{c_P}$ instead of $\frac{X_P}{c_P}$ (here X_P is simply the range or Lebesgue measure of the values of the random variable X in the part P), because the part P need not be uniformly distributed in general, and hence comparing the ratio $\frac{X_P}{c_P}$ would not amount to an even comparison across parts. The value of D_P however, is an equivalent uniform representation of the part P which maintains the same conditional entropy of the part P. In other words, by re-interpreting the part P as an equivalent equi-probable distribution that maintains the entropic uncertainty, the range of the re-interpreted uniform distribution (which is precisely D_P) gives us an even playing field to compute and compare the degree of uniformity of the part P. The ratio $\frac{D_P}{c_P}$ therefore, evaluates the extent of uniformity (in an entropically equivalent sense) per unit frequency that originates from the part P.

Hence, in this way, given two parts P_1 and P_2, we can directly compare the ratios $\frac{D_{P_1}}{c_{P_1}}$ and $\frac{D_{P_2}}{c_{P_2}}$ on an even playing field where we have redrawn the parts P_1 and P_2 into their SEE equivalents which are both uniformly distributed, even though the original parts P_1 and P_2 are not. This justifies the coinage of the term *degree of uniformity* of the part P for the ratio $\frac{D_P}{c_P}$. We summarize the mechanics of the quantification of the degree of uniformity/inequality using the figure 1.

3.2. $\frac{D_P}{c_P}$ as degree of inequality

Next, we illustrate the alternate viewpoint of *degree of inequality* of the part P for the ratio $\frac{D_P}{c_P}$. For both a discrete uniform distribution with N equi-probable types, or a continuous uniform distribution supported on a finite interval (a,b), the standard deviation can be shown to be of the order of N or $(b - a)$ respectively.

By definition, the value of D_P for a given part P is mathematically equal to N or $(b - a)$ (in its Shannon Equivalent Equi-probable form for the part P) in the discrete or continuous uniform distributions described

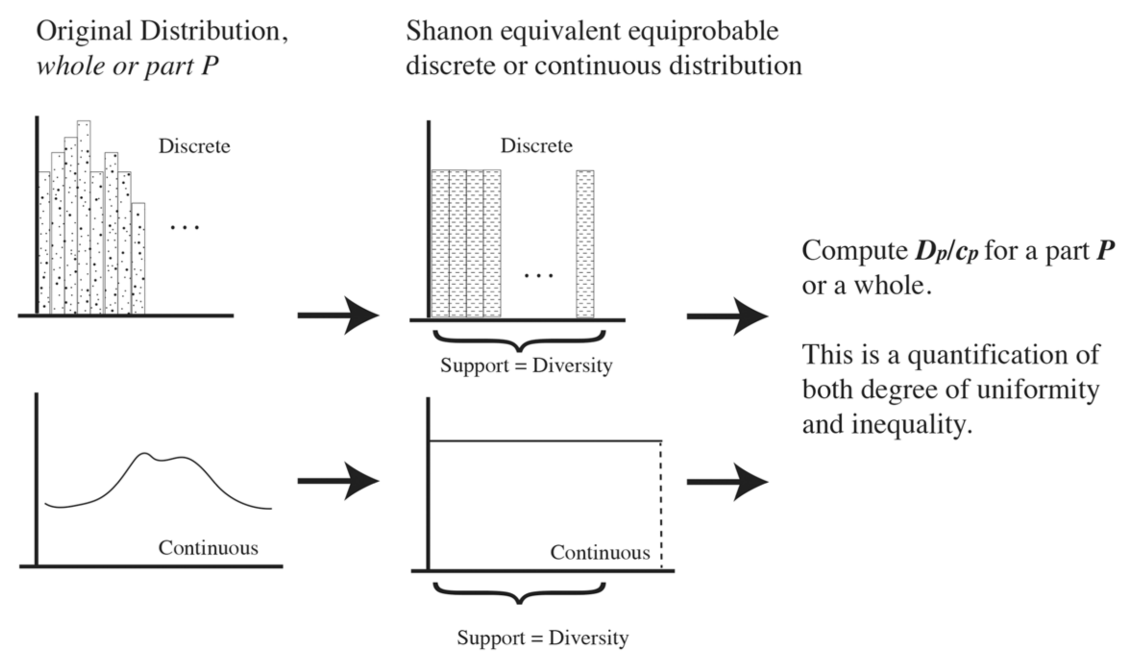

Figure 1. Picture showing the mechanics of computing the degree of uniformity or inequality.

here. In other words, D_P is mathematically equivalent to the standard deviation or spread of the part P in its Shannon Equivalent Equi-probable form. Hence for an SEE part with diversity D_P (which is nothing but the support of the SEE part), the value of D_P is actually a measure of the spread of the original random variable X_P over the cumulative probability c_P for the part P, but in a Shannon equivalent equi-probable way. Hence, a part P that has a larger ratio of $\frac{D_P}{c_P}$ has a larger amount (or spread) of the random variable (albeit in its SEE form) spread out over a smaller frequency or cumulative probability.

This in turn means that there is a larger localization of inequality of the random variable over the part P compared to other parts that have a smaller value for the ratio $\frac{D_P}{c_P}$. Comparing the ratio $\frac{D_P}{c_P}$ in this way for different parts is a scale free way to compare the localization of inequality in parts because the variation of the original part P is converted into an equivalent equi-probable variation and we are calculating this SEE variation per unit cumulative frequency.

In a sense, we are using the entropy of the part P (its exponential actually) and its cumulative frequency (which is scale free since it varies from 0 to 1 for all distributions) to make comparisons. In summary, the ratio $\frac{D_P}{c_P}$ computes the Shannon equivalent equiprobable spread per unit cumulative frequency for the part P in a scale free manner and hence, the ratio $\frac{D_P}{c_P}$ has the interpretation of *degree of inequality* as well.

4. Power law distribution

The distribution of a random variable X is said to follow the power law if the probability of measuring it is inversely proportional to a fixed power of the quantity itself. Power law distributions are also referred to as Pareto or Zipf distributions depending on the literature. These distributions occur naturally in a variety of fields such as physics, earth sciences, biology, computer science, sociology, and others. Due to the decreasing nature of the probability density $p_z(x)$ (we use the subscript letter z for Zipf) of a power law distribution, stemming from the inverse power-relationship with x, such distributions are inherently right-skewed, i.e. larger values of x are less probable. At the outset, we define a power law distribution below:

Definition 4.1. A continuous random variable X is said to follow a power law distribution if its probability density function denoted by $p_z(x)$ satisfies the the following:

$$p_z(x) = \frac{\alpha - 1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha}; \quad \text{where } \alpha > 1; \quad x \in (x_{\min}, \infty) \tag{7}$$

$$= Cx^{-\alpha}; \quad \text{where } C = (\alpha - 1)x_{\min}^{\alpha-1}. \tag{8}$$

In the above definition, the constant C is normalized to make the total probability equal to 1. Also, α is typically chosen to be larger than 1 for the mean to be well defined, and $x_{\min} > 0$ is a positive minimum value of the random variable to avoid singularities in the distribution. In this section, we state and prove some preliminary results about the power law distribution defined in definition 4.1. We recall the following two theorems from (Rajaram *et al* 2024).

Theorem 4.1. *Given a power law distribution as in definition 4.1, its entropy is given by*

$$H = \ln\left(\frac{x_{\min}}{(\alpha - 1)}\right) + \frac{\alpha}{(\alpha - 1)}, \ \alpha > 1. \tag{9}$$

We denote the total diversity of the power law distribution $D_{(x_{\min},\infty)}$ by the symbol D just for simplicity.

We can easily show that

$$\begin{aligned} c_{(x_{\min},x)} &= C\int_{x_{\min}}^{x} t^{-\alpha}dt \\ &= 1 - \left(\frac{x}{x_{\min}}\right)^{(1-\alpha)}. \end{aligned}$$

Next, we recall a theorem from (Rajaram *et al* 2024) that calculates an explicit formula for the slope of diversity curve of the power law distribution.

Theorem 4.2. *Given a power law distribution as in definition 4.1, the slope of diversity curve which plots $c_{(x_{\min},x)}$ on the x-axis and $c_{(x_{\min},x)} * \ln(A_{(x_{\min},x)})$ on the y-axis has the following explicit formula:*

$$c_{(x_{\min},x)}\ln\left(A_{(x_{\min},x)}\right) = \frac{-\alpha}{(-\alpha + 1)}(1 - c_{(x_{\min},x)})\ln\left(1 - c_{(x_{\min},x)}\right). \tag{10}$$

Also, the slope of the tangent s_x of the slope of diversity curve at $c_{(x_{\min},x)}$ is given by:

$$\begin{aligned} s_x &= \frac{\alpha}{\alpha - 1)}\{-\ln(1 - c_{(x_{\min},x)}) - 1\} \\ &= \alpha\ln\left(\frac{x}{x_{\min}}\right) - \frac{\alpha}{(\alpha - 1)}. \end{aligned}$$

4.1. Case-based entropy curve for a power law distribution

We next compute an explicit formula for the case-based entropy curve of the power law distribution.

Theorem 4.3. *Given a power law distribution as in definition 4.1, the case-based entropy curve which plots $c_{(x_{\min},x)}$ on the x-axis and $C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D}$ on the y-axis has the following explicit formula:*

$$C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D} = c_{(x_{\min},x)}\left(1 - c_{(x_{\min},x)}\right)^{\frac{\alpha}{(\alpha-1)}\frac{(1-c_{(x_{\min},x)})}{c_{(x_{\min},x)}}} \tag{11}$$

Proof. The proof is a calculation.

$$\begin{aligned} c_{(x_{\min},x)}\ln(A_{(x_{\min},x)}) &= \frac{\alpha}{(\alpha - 1)}(1 - c_{(x_{\min},x)})\ln(1 - c_{(x_{\min},x)}) \\ \ln(A_{(x_{\min},x)}) &= \frac{\alpha}{(\alpha - 1)}\frac{(1 - c_{(x_{\min},x)})}{c_{(x_{\min},x)}}\ln(1 - c_{(x_{\min},x)}) \\ \frac{D_{(x_{\min},x)}}{c_{(x_{\min},x)}D} &= A_{(\min,x)} = (1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}\frac{(1-c_{(x_{\min},x)})}{c_{(x_{\min},x)}}} \\ C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D} &= c_{(x_{\min},x)}(1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}\frac{(1-c_{(x_{\min},x)})}{c_{(x_{\min},x)}}} \end{aligned}$$

This proves the Theorem.

Next, we show that the case-based entropy curve for the power law distribution is an increasing function.

Theorem 4.4. *Given a power law distribution as in definition 4.1, the case-based entropy curve, which plots $c_{(x_{\min},x)}$ on the x-axis and $C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D}$ on the y-axis, is an increasing function of $c_{(x_{\min},x)}$.*

Proof. For simplification of notation, let $y = g(x) = C_{(x_{\min},x)}$ and $x = c_{(x_{\min},x)}$ and note that $y > 0$. Then we have

$$y = g(x) = x(1-x)^{\frac{\alpha}{(\alpha-1)}\left(\frac{1-x}{x}\right)}.$$

Let's compute the first derivative, with respect to x using the logarithmic differentiation technique.

$$\ln(y) = \frac{\alpha}{(\alpha-1)}\left(\frac{1-x}{x}\right)\ln(1-x) + \ln(x)$$

$$\frac{y'}{y} = \frac{\alpha}{(\alpha-1)}\left\{\ln(1-x)\left[\frac{x(-1)-(1-x)}{x^2}\right] + \frac{(1-x)}{x}\left(\frac{-1}{(1-x)}\right)\right\}$$

$$\frac{y'}{y} = \frac{\alpha}{(\alpha-1)}\left\{\frac{-\ln(1-x)}{x^2} - \frac{1}{x}\right\} + \frac{1}{x}$$

$$\frac{y'}{y} = \frac{\alpha}{(\alpha-1)}\left\{\frac{-\ln(1-x)-x}{x^2}\right\} + \frac{1}{x}$$

Define $f(x)$ as follows:

$$f(x) = -\ln(1-x) - x, \quad \text{noting that} \quad f(0) = 0 \quad \text{and}$$

$$f'(x) = \frac{1}{1-x} - 1 = \frac{x}{1-x} > 0, \quad \text{if } 0 < x < 1.$$

Hence $f(x) > f(0) = 0$ if $0 < x < 1$.

So, this directly implies that $y' = g'(x) > 0$, when $\alpha > 1$ and $0 < x < 1$, or y is an increasing function of x. This proves the Theorem.

Next, we show that the $60 - 40$ rule that was empirically observed for power law distributions in (Castellani and Rajaram 2016) is analytically true for power laws. We define the $60 - 40$ rule below first.

Definition 4.2. Given a power law distribution as in definition 4.1, its case-based entropy curve, which plots $c_{(x_{\min},x)}$ on the x-axis and $C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D}$ on the y-axis, satisfies the $60 - 40$ rule if the curve passes through the following region:

$$\{0.6 < c_{(x_{\min},x)} < 1,\ 0 < C_{(x_{\min},x)} < 0.4\}.$$

Remark 4.1. Definition 4.2 means that at least 60% of the cases are situated in the first 40% or less of Shannon Equivalent Equi-probable (SEE) part of the distribution. The $60 - 40$ rule demonstrates the severity of restriction of diversity in power laws (and in general right-tailed distributions as empirically observed in (Castellani and Rajaram 2016)). It says that at least 60% of the cases in all power laws are restricted to the first 40% or less of Shannon Equivalent Equi-probable types. This means that a majority of the distribution i.e., 60% or more is situated in the first 40% or less of the diversity (or uniformity) of the distribution in an equivalent equi-probable way.

Theorem 4.5. *Given a power law distribution as in definition 4.1, its case-based entropy curve which plots $c_{(x_{\min},x)}$ on the x-axis and $C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D}$ on the y-axis, satisfies the $60 - 40$ rule as stated in definition 4.2.*

Proof. In theorem 4.4 we showed that the case-based entropy curve is an increasing function of $c_{(x_{\min},x)}$. Using the same simplification of notation, let $y = g(x) = C_{(x_{\min},x)}$ and $x = c_{(x_{\min},x)}$. Since $g(x)$ is increasing, we know that

$$g(x) > g(0.6) = 0.326^{\left(\frac{\alpha}{\alpha-1}\right)}, \quad \text{for} \quad 0.6 < x < 1.$$

As α approaches infinity, $\frac{\alpha}{\alpha-1}$ decreases to 1, and

$$g(0.6) = 0.326^{\left(\frac{\alpha}{\alpha-1}\right)} \uparrow 0.326.$$

Hence the maximum limiting value of $g(0.6) = 0.326$ for any $\alpha \to \infty$, with $g(0.6) < 0.326$, for $1 < \alpha < +\infty$.

Since $g(x)$ is increasing and $g(0.6) = 0.326$ in the maximum $\alpha-$ limit, this means that

$$g(x) = x(1-x)^{\left(\frac{1-x}{x}\right)\cdot\frac{\alpha}{(\alpha-1)}} \tag{12}$$

passes through the region $\{0.6 < x < 1 \quad \text{and} \quad 0 < y < 0.40, \quad \text{actually} \quad 0.326\}$.

Thus, we have proven the '60/40 Rule'. That is, at least 60% or more of cases $(0.6 < c_{(x_{\min,x})} < 1)$ are within the first 40% of equiprobable types $(0 < C_{(x_{\min,x})} < 0.4)$.

Figure 2 shows the case-based entropy curve for the power law distribution for various choices of α:

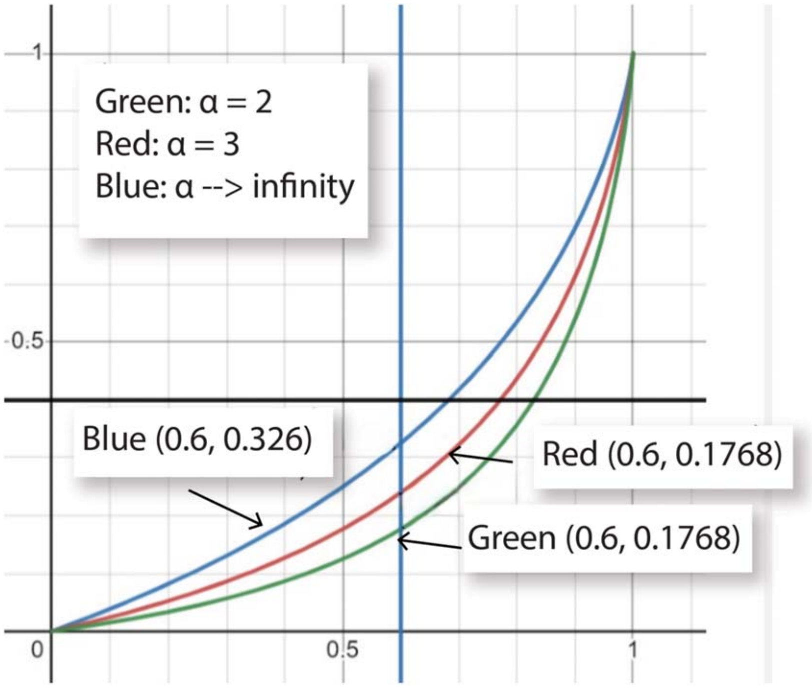

Figure 2. Case-based entropy curve for the power law distribution with $c_{(x_{\min},x)}$ on the x-axis and $C_{(x_{\min},x)}$ on the y-axis for various choices of α showing that it satisfies the $60-40$ rule.

In fact, from equation (12) we have that

$$g(x) = x(1-x)^{\left(\frac{1-x}{x}\right)\cdot\frac{\alpha}{(\alpha-1)}} \uparrow x(1-x)^{\left(\frac{1-x}{x}\right)} \text{as} \alpha \uparrow \infty.$$

Hence we have the following more general definition and theorem for which theorem 4.5 is a special case when $x = 0.6$.

Definition 4.3. Given a power law distribution as in definition 4.1, its case-based entropy curve which plots $c_{(x_{\min},x)}$ on the x-axis and $C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D}$ on the y-axis, satisfies a $c - g(c)$ rule if the curve passes through the following region:

$$\{c < c_{(x_{\min},x)} < 1,\ \ 0 < C_{(x_{\min},x)} < g(c)\}.$$

Theorem 4.6. *Given a power law distribution as in definition 4.1, the case-based entropy curve, which plots $c_{(x_{\min},x)}$ on the* x*-axis and $C_{(x_{\min},x)} = \frac{D_{(x_{\min},x)}}{D}$ on the* y*-axis, satisfies the $c - g(c)$ rule as stated in definition 4.3 with $g(c) = c(1-c)^{\left(\frac{1-c}{c}\right)}$.*

Proof. The proof follows the same steps as in theorem 4.5.

In summary, we have analytically proved the $60-40$ rule for the power law distribution in this section that was empirically observed in (Castellani and Rajaram 2016) for a variety of systems that exhibited a right tailed distribution. We have also generalized the $60-40$ rule to a $c - g(c)$ rule which equates to the $60-40$ rule when c is chosen to be 0.6.

5. Comparing degree of uniformity of parts of the power law distribution

The $60-40$ rule in the previous section suggests that less uniformity is coming from a majority of the distribution since 60% or more of the distribution is in the first 40% or less of uniformity. This is a first step to quantifying the inequality in power laws, since it is already known that the tail is more uniformly distributed compared to the initial part. This also begs the question of quantification of the ratio of degree of uniformity of the right part of the power law to the left part, or more generally between any two parts as well if needed. This is precisely what we do in this section.

In this section, we compare the degree of uniformity (or inequality) of parts of the power law using the explicit expression for the slope of diversity curve that was derived in theorem 4.2.

Theorem 5.1. *Given a power law distribution as in definition 4.1, the ratio of degree of uniformity of the parts* $(x_{\min}, x)$ *and* (x, ∞) *is given by the following:*

$$\frac{A_{(x,\infty)}}{A_{(x_{\min},x)}} = (1 - c_{(x_{\min},x)})^{-\frac{\alpha}{(\alpha-1)}\left(\frac{1}{c_{(x_{\min},x)}}\right)}. \tag{13}$$

Proof. Recall from theorem 4.2 that

$$A_{(x_{\min},x)} = (1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}\frac{(1-c_{(x_{\min},x)})}{c_{(x_{\min},x)}}}.$$

We also have

$$A_{(x_{\min},x)}^{c_{(x_{\min},x)}} \cdot A_{(x,\infty)}^{(1-c_{(x_{\min},x)})} = 1.$$

So,

$$\begin{aligned}
(1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}(1-c_{(x_{\min},x)})} A_{(x,\infty)}^{(1-c_{(x_{\min},\infty)})} &= 1\\
(1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}} A_{(x,\infty)} &= 1\\
A_{(x,\infty)} &= (1 - c_{(x_{\min},x)})^{-\frac{\alpha}{(\alpha-1)}}.
\end{aligned}$$

Taking ratios, we have

$$\begin{aligned}
\frac{A_{(x,\infty)}}{A_{(x_{\min},x)}} &= \frac{(1 - c_{(x_{\min},x)})^{-\frac{\alpha}{(\alpha-1)}}}{(1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}\left(\frac{(1-c_{(x_{\min},x)})}{c_{(x_{\min},x)}}\right)}}\\
&= (1 - c_{(x_{\min},x)})^{-\frac{\alpha}{(\alpha-1)}\left(\frac{1}{c_{(x_{\min},x)}}\right)}.
\end{aligned}$$

This proves the Theorem.

Remark 5.1. We remark that the following calculation holds to compute $\tilde{x}$ so that $c_{(x_{\min},\tilde{x})} = c_{(\tilde{x},\infty)} = 0.5$.

$$\begin{aligned}
c_{(x_{\min},\tilde{x})} = 1 - \left(\frac{\tilde{x}}{x_{\min}}\right)^{(-\alpha+1)} &= 0.5 \quad \text{so}\\
\left(\frac{\tilde{x}}{x_{\min}}\right)^{(-\alpha+1)} &= 0.5,\\
\left(\frac{x_{\min}}{\tilde{x}}\right)^{(\alpha-1)} &= 0.5,\\
x_{\min} &= \frac{\tilde{x}}{2^{\frac{1}{(\alpha-1)}}}, \quad \text{or}\\
\tilde{x} &= x_{\min} \cdot 2^{\frac{1}{(\alpha-1)}}.
\end{aligned}$$

For such an $\tilde{x}$, substituting in equation (13)

$$A_{(x_{\min},\tilde{x})} = (0.5)^{\frac{\alpha}{(\alpha-1)}}, \tag{14}$$

and

$$A_{(\tilde{x},\infty)} = 2^{\frac{\alpha}{(\alpha-1)}}. \tag{15}$$

The first equation provides the ratio of the degree of uniformity of the interval $(x_{\min}, \tilde{x})$ to the entire distribution $(x_{\min}, \infty)$. The second equality provides the ratio of the degree of uniformity of the interval $(\tilde{x}, \infty)$ to the entire distribution $(x_{\min}, \infty)$.

Finally,

$$\frac{A_{(\tilde{x},\infty)}}{A_{(x_{\min},\tilde{x})}} = \frac{D_{(\tilde{x},\infty)}/c_{(\tilde{x},\infty)}}{D_{(x_{\min},\tilde{x})}/c_{(x_{\min},\tilde{x})}} = 4^{\frac{\alpha}{(\alpha-1)}}. \tag{16}$$

The above observations lead to the following theorem.

Theorem 5.2. *Let* $\tilde{x}$ *be the halfway point in the power law distribution as in remark 5.1 for a power law distribution, from definition 4.1. Then we have the following.*

1. *The first 50% of the distribution from* $(x_{\min}, \tilde{x})$ *is at most 50% as uniformly distributed as the entire distribution* $(x_{\min}, \infty)$.
2. *The second half* $(\tilde{x}, \infty)$ *is at least 200% (or twice) as uniformly distributed as the entire distribution* $(x_{\min}, \infty)$.

3. *The second half* $(\tilde{x}, \infty)$ *is at least 4 times as uniformly distributed as the first half* $(x_{\min}, \tilde{x})$

Proof.

(1) From equation (14), as $\alpha \to \infty$, $A_{(x_{\min},\tilde{x})}$ increases to 0.5 and $A_{(\tilde{x},\infty)}$ decreases to 2. This means that the first 50% of the distribution from $(x_{\min}, \tilde{x})$ is at most 50% as uniformly distributed as the entire distribution $(x_{\min}, \infty)$.

(2) Also, from equation (15) the second half $(\tilde{x}, \infty)$ is at least 200% (or twice) as uniformly distributed as the entire distribution $(x_{\min}, \infty)$.

(3) From equation (16), the second half $(\tilde{x}, \infty)$ is at least 4 times as uniformly distributed as the first half $(x_{\min}, \tilde{x})$ since $\frac{A_{(\tilde{x},\infty)}}{A_{(x_{\min},\tilde{x})}}$ decreases to 4 as $\alpha \to +\infty$.

This proves the Theorem.

Remark 5.2. We can generalize the results above for the right p% of the power law, i.e. for the choice of $1 - c_{(x_{\min},x)} = p$. Thus,

$$c_{(x_{\min},x)} = 1 - \left(\frac{x}{x_{\min}}\right)^{(-\alpha+1)} \Rightarrow x = \frac{x_{\min}}{p^{(\alpha-1)}}$$

.For such an x, we have

$$A_{(x_{\min},x)} = (1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}\frac{(1-c_{(x_{\min},x)})}{c_{(x_{\min},x)}}} = p^{\frac{\alpha p}{(\alpha-1)(1-p)}},$$

$$A_{(x,\infty)} = (1 - c_{(x_{\min},x)})^{-\frac{\alpha}{(\alpha-1)}} = p^{-\frac{\alpha}{\alpha-1}}, \quad \text{and}$$

$$\frac{A_{(x,\infty)}}{A_{(x_{\min},x)}} = p^{-\frac{\alpha}{(\alpha-1)}\left(\frac{1}{1-p}\right)}.$$

Hence, we have the following more general theorem from the above observations for the power law by taking similar limits as $\alpha \to \infty$:

Theorem 5.3. *Let x be the* $(1 - p) - th$ *percentile in the power law distribution as in remark 5.2 for a power law distribution as in definition 4.1. Then we have the following.*

1. *The first* $(1 - p)$% *of the distribution from* $(x_{\min}, x)$ *is at most* $p^{\frac{p}{(1-p)}}$% *as uniformly distributed as the entire distribution* $(x_{\min}, \infty)$.
2. *The last* p% *of the distribution* (x, ∞) *is at least* $\frac{1}{p}$% *as uniformly distributed as the entire distribution* $(x_{\min}, \infty)$.
3. *The last* p% *of the distribution* (x, ∞) *is at least* $p^{-\left(\frac{1}{1-p}\right)}$% *as uniformly distributed as the first* $(1 - p)$% *of the distribution* $(x_{\min}, x)$.

Proof. The proof follows the same steps as in theorem 5.2.

6. Entropic comparison of power law and exponential distribution

In this section, we demonstrate that we can compare two different distributions with possibly completely different scales of variation for the random variable X and the associated probability densities. This is possible by studying the variation of degree of uniformity (or inequality) of parts through the ratio $\frac{D_P}{c_P}$ as a function of cumulative probability c using the slope of diversity curve. This is because the cumulative probability c always varies from 0 to 1 and this provides a *scale free* way to look at the parts of the original random variable. In addition, the diversity D_P is the support of an abstract SEE (Shannon Equivalent Equi-probable) distribution which is uniformly distributed, and hence gives us a way to evenly compare parts even though the variation of probabilities of said parts might be wildly different from each other and from the uniform distribution.

Consider the exponential distribution as defined below.

Definition 6.1. A continuous random variable X is said to follow an exponential distribution if its probability density function and cumulative probability distribution denoted by $p_e(x)$ and $c_{(0,x)}$ respectively, satisfy the the following:

$$p_e(x) = \lambda e^{-\lambda x} \quad \text{and} \quad c_{(0,x)} = 1 - e^{-\lambda x}, \quad x \in (0, \infty). \tag{17}$$

We recall the following theorem from (Rajaram *et al* 2024).

Theorem 6.1. *Given an exponential distribution as in definition 6.1, the slope of diversity curve and the slope of tangent for the same, are given as follows:*

$$c_{(0,x)} \ln(A^e_{(0,x)}) = -\lambda x e^{-\lambda x} \text{ and} \tag{18}$$

$$s_x = \lambda x - 1 = -\ln(1 - c_{(0,x)}) - 1. \tag{19}$$

We next derive an expression for the ratio of degree of uniformity of parts $(0, x)$ and (x, ∞) to the whole $(0, \infty)$ for an exponential distribution for arbitrary choice of $x \in (0, \infty)$.

Theorem 6.2. *Given an exponential distribution as in definition 6.1, the ratio of degree of uniformity of parts $(0, x)$ and (x, ∞) respectively with respect to the whole $(0, \infty)$ denoted by $A^e_{(0,x)}$ and $A^e_(x, \infty)$ respectively, are as follows.*

$$A^e_{(0,x)} = (1 - c_{(0,x)})^{(1-c_{(0,x)})/c_{(0,x)}} \tag{20}$$

$$A^e_{(x,\infty)} = \frac{1}{(1 - c_{(0,x)})} \tag{21}$$

Proof. We know from theorem 6.1 that

$$c_{(0,x)} \ln(A^e_{(0,x)}) = -\lambda x e^{-\lambda x}.$$

Also,

$$\begin{aligned} c_{(0,x)} &= 1 - e^{-\lambda x} \\ e^{-\lambda x} &= 1 - c_{(0,x)} \\ -\lambda x &= \ln(1 - c_{(0,x)}). \end{aligned}$$

So,

$$c_{(0,x)} \ln(A^e_{(0,x)}) = (1 - c_{(0,x)}) \ln(1 - c_{(0,x)}),$$

and

$$\begin{aligned} A^e_{(0,x)} &= (1 - c_{(0,x)})^{(1-c_{(0,x)})/c_{(0,x)}} \\ {A^e_{(0,x)}}^{c_{(0,x)}} \cdot {A^e_{(x,\infty)}}^{(1-c_{(0,x)})} &= 1 \\ (1 - c_{(0,x)})^{(1-c_{(0,x)})/c_{(0,x)}} \cdot {A^e_{(x,\infty)}}^{(1-c_{(0,x)})} &= 1 \\ A^e_{(x,\infty)} &= \frac{1}{(1 - c_{(0,x)})}. \end{aligned}$$

Remark 6.1. Now, we choose $\tilde{x}$ so that $c_{(0,\tilde{x})} = 0.5$. Using a bit of algebra, this occurs when $\tilde{x} = \frac{\ln(2)}{\lambda}$.
Then,

$$A^e_{(0,\tilde{x})} = (1 - 0.5)^{\left(\frac{1-0.5}{0.5}\right)} = 0.5,$$

and

$$A^e_{(\tilde{x},\infty)} = \frac{1}{(1 - 0.5)} = 2.$$

Interestingly enough, for the exponential distribution, the first half is exactly 50% as uniform as the entire distribution and the second half is exactly twice as uniform as the entire distribution. Also, the second half is 4 times as uniform as the first half. More generally, the first $(1 - p)$% of the exponential distribution is exactly $p^{\frac{p}{1-p}}$% as uniform as the entire distribution and the last p% is exactly $\frac{1}{p}$% as uniform as the entire distribution. Finally, the last p% of the distribution is exactly $p^{-\frac{1}{1-p}}$% as uniformly distributed as the first $(1 - p)$% of the exponential distribution.

We next establish an important and interesting relationship between the power law and the exponential distributions. In what follows, a subscript or superscript of z refers to the power law and e refers to the exponential distribution respectively.

Theorem 6.3. *Let $p_z(x)$ and $p_e(x)$ be the probability density functions for the power law and the exponential distributions respectively, as defined in definitions 4.1 and 6.1. Let $A^z_{(x_{\min},x)}$ and $A^e_{(0,x)}$ be the ratio of degrees of uniformity of the power law and exponential distributions respectively, each with their corresponding total diversities D_z and D_e. Then the following is true:*

$$\lim_{\alpha\to\infty} A^z_{(x_{\min},x)} = A^e_{(0,x)} \tag{22}$$

$$\lim_{\alpha\to\infty} A^z_{(x,\infty)} = A^e_{(x,\infty)}. \tag{23}$$

Proof. For general $x \in (x_{\min}, \infty)$ and $x \in (0, \infty)$ respectively, for the power law and the exponential distributions, we have the following:

$$A^z_{(x_{\min},x)} = (1 - c_{(x_{\min},x)})^{\frac{\alpha}{(\alpha-1)}\frac{(1-c_{(x_{\min},x)})}{c_{(x_{\min},x)}}},$$

$$A^z_{(x,\infty)} = (1 - c_{(x_{\min},x)})^{-\frac{\alpha}{(\alpha-1)}},$$

$$A^e_{(0,x)} = (1 - c_{(0,x)})^{\frac{(1-c_{(0,x)})}{c_{(0,x)}}}, \quad \text{and}$$

$$A^e_{(x,\infty)} = \frac{1}{(1 - c_{(0,x)})}.$$

The result follows by taking the limit as $\alpha \to \infty$.

Remark 6.2. The key point here is that $c_{(x_{\min},x)}$ and $c_{(0,x)}$ are both cumulative probabilities from 0 to 1 for the power law and the exponential distributions respectively, and we don't distinguish between the cumulative probability variable for the power law and that for the exponential distribution. In this case, we are primarily interested in the variation of mathematical diversity of the two distributions as a function of their cumulative probabilities. The degree of uniformity is the ratio $\frac{D_P}{c_P}$, where D_P is the support of the SEE distribution of the part that is uniformly distributed. Hence, the original variation of probabilities in the parts are redrawn as equivalent uniform distributions thereby providing an even playing field of comparison. In this sense, we can actually forget about the original random variable once we have A as a function of c. Hence, from the standpoint of slope of diversity, the exponential distribution is the entropic limit of the power law distribution as $\alpha \to \infty$. Such a comparison is inherently *scale free* since the range or variation of the original distributions (power law and exponential) are subsumed by the cumulative probability random variable c which varies from 0 to 1 in both cases.

To our knowledge, this connection between the exponential distribution and power law has not been discovered before.

Remark 6.3. Let the subscript (or superscript) z denote the power law and e denote the exponential distribution, and let $\tilde{x}$ be such that $c_{(x_{\min},\tilde{x}_z)} = c_{(0,\tilde{x}_e)} = 0.5$, where $\tilde{x}_z = x_{\min} \cdot 2^{(\alpha-1)}$ and $\tilde{x}_e = \ln(2)/\lambda$, i.e. $\tilde{x}_z$ and $\tilde{x}_e$ divide the power law and the exponential distributions, respectively, into two equal halves.

Then

$$A^z_{(x_{\min},\tilde{x}_z)} = 0.5^{\left(\frac{\alpha}{\alpha-1}\right)}$$

and

$$A^e_{(0,\tilde{x}_e)} = 0.5.$$

Hence,

$$\lim_{\alpha\to\infty} A^z_{(x_{\min},\tilde{x}_z)} = A^e_{(0,\tilde{x}_e)}.$$

The slope of diversity of the power law is

$$c_{(x_{\min},x)} \ln (A^z_{(x_{\min},x)}) = \left(\frac{\alpha}{\alpha - 1}\right)(1 - c_{(x_{\min},x)}) \ln (1 - c_{(x_{\min},x)}))$$

The slope of diversity of the exponential distribution is

$$c_{(0,x)} \ln (A^e_{(0,x)}) = (1 - c_{(0,x)})\ln(1 - c_{(0,x)}).$$

Clearly,

$$\lim_{\alpha\to\infty} c_{(x_{\min},x)} \ln (A^z_{(x_{\min},x)}) = c_{(0,x)} \ln (A^e_{(0,x)}),$$

where the x in $A^z_{(x_{\min},x)}$ and $A^e_{(0,x)}$ refer to the random variable in the power law and the exponential distributions, respectively.

Also,

$$s_x^z = \left(\frac{\alpha}{\alpha - 1}\right)(-\ln(1 - c_{(x_{\min},x)})) - 1)$$

and

$$s_x^e = -\ln(1 - c_{(0,x)}) - 1.$$

Hence,

$$\lim_{\alpha \to \infty} s_x^z = s_x^e.$$

Our conclusion is that the exponential distribution is the entropic limit of power law distributions as $\alpha \to \infty$.

Note: This does not mean that the densities p_z and p_e follow this limit. The reason for this is that the diversities D_z and D_e are different, hence $p_z(x) = e^{-s_x^z}/D_z$ and $p_e(x) = e^{-s_x^e}/D_e$ are different. This also does not contradict the equivalence theorem proved in (Rajaram *et al* 2023, 2024) between the original probability distribution and the slope of diversity curve, since the probability spaces for the exponential distribution $(0, \infty)$ and the power law distribution $(x_{\min}, \infty)$ can never be the same. This is because the power has to necessarily start from $x_{\min} > 0$.

Remark 6.4. We remark again, to emphasize that our focus in the comparison of the power law and the exponential distributions is from the lens of mathematical diversity and its variation as a function of their respective cumulative probabilities. Since the cumulative probabilities for both distributions range from 0 to 1, this gives us a way to compare and contrast several things such as the diversity of parts, the degree of uniformity of parts etc., for parts across the two distributions, without paying heed to the actual differences in the variation of the probability itself as captured by their respective densities. We refer to such comparisons as *scale free* since the underlying variable is the cumulative probability *c*, which varies between 0 and 1 for all probability distributions. We also note that we are able to obtain relationships (in the case of the power law and the exponential distributions, this was a limiting relationship) between the slope of diversity curves and the slope of tangent of said curves as well. Mathematically speaking, diversity is the exponential of entropy, and functional relationships between the diversity of parts within a given distribution or across different distributions (such as power law and exponential) are equivalently, statements of relationships between the variation of entropic uncertainty as a function of their respective cumulative probabilities. For this reason, we have termed the limiting relationship between the exponential and the power law distributions in the paper as *entropic limits*. We have used the case-based entropy and the slope of diversity curve as vehicles to explore such relationships between mathematical diversity and the cumulative probability of parts both within a given distribution and across different distributions. Hence, in a sense, we have devised a very general mathematical machinery to compare the parts of a given distribution and also across different distributions in a scale-free way from the standpoint of mathematical diversity.

7. Geometric distribution

Although the previous examples were chosen to be continuous probability distributions due to their importance, the same ideas can be extended to discrete distributions as well. We demonstrate our ideas using the geometric distribution as the third example.

Definition 7.1. A discrete random variable X is said to follow the geometric distribution if its probabilities p_i satisfy the following:

$$p_i = pq^{(i-1)}, \quad q = 1 - p, \quad \text{for} \quad i = 1, 2, 3, \ldots. \tag{24}$$

We recall some of the important formulas for the geometric distribution below. D stands for the diversity of the entire distribution, D_k stands for the diversity of the part $\{1, k\}$ until the index k, c_k stands for the cumulative probability up to the index k, $\hat{p}_i$ stands for the normalized probability for the part $\{1, k\}$ and p, q are the parameters of the geometric distribution denoting success and failure probabilities. Let

$$
\begin{aligned}
D &= \frac{1}{p \cdot q^{(q/p)}} \\
\hat{p}_i &= \frac{p \cdot q^{(i-1)}}{(1 - q^i)} \quad \text{for } i = 1, 2, 3, \ldots, k; \quad \text{and} \quad c_k = 1 - q^k \\
D_k &= D(q^{(k-1)q^k/(1-q^k)} \cdot \underbrace{(1 - q^k)}_{\text{This is } c_k}) \\
A_{\{1,k\}} &= \frac{D_k}{Dc_k} = q^{(k-1)q^k/(1-q^k)}.
\end{aligned}
$$

Theorem 7.1. *Let $\hat{k}$ denote the index corresponding to the $(1 - t)$-th percentile for a geometric distribution as in definition 7.1. Then the ratio of degree of uniformity of the part $\{1, \hat{k}\}$ and the part $\{\hat{k}, \infty\}$ is given by the following:*

$$A_{\{\hat{k},\infty\}}/A_{\{1,\hat{k}\}} = t^{-\left(1-\frac{\ln q}{\ln t}\right)\cdot\left(\frac{1}{1-t}\right)}. \tag{25}$$

Proof. Noting that $q = (1 - c_k)^{1/k}$ and $q^k = 1 - c_k$, we have the following:

$$A_{\{1,k\}} = (1 - c_k)^{\left(\frac{k-1}{k}\right)\cdot\left(\frac{1-c_k}{c_k}\right)}, \quad \text{or} \tag{26}$$

$$A_c = (1 - c)^{\left(\frac{k-1}{k}\right)\cdot\left(\frac{1-c}{c}\right)} \tag{27}$$

Since

$$A_{\{1,k\}}^{c_k} \cdot A_{\{k,\infty\}}^{(1-c_k)} = 1,$$

we have

$$
\begin{aligned}
A_{\{k,\infty\}}^{(1-c_k)} &= (1 - c_k)^{-((k-1)/k)(1-c_k)} \\
A_{\{k,\infty\}} &= (1 - c_k)^{-((k-1)/k)},
\end{aligned}
$$

and so

$$
\begin{aligned}
A_{\{k,\infty\}}/A_{\{1,k\}} &= \frac{(1 - c_k)^{-((k-1)/k)}}{(1 - c_k)^{-((k-1)/k)((1-c_k)/c_k)}} \\
&= (1 - c_k)^{-\frac{(k-1)}{k}\cdot\frac{1}{c_k}}.
\end{aligned}
$$

Let $\hat{k}$ be the index corresponding to the right t%. Then,

$$
\begin{aligned}
1 - c_{\{1,\hat{k}\}} &= q^{\hat{k}} = t, \\
\hat{k} &= \frac{\ln(t)}{\ln(q)}, \quad \text{and} \\
\frac{\hat{k} - 1}{\hat{k}} &= \frac{\ln(t) - \ln(q)}{\ln(t)}.
\end{aligned}
$$

For such a $\hat{k}$, we have

$$
\begin{aligned}
A_{\{1,\hat{k}\}} &= (1 - c_{\{1,\hat{k}\}})^{\frac{(\hat{k}-1)}{\hat{k}}\cdot\frac{(1-c_{\{1,\hat{k}\}})}{c_{\{1,\hat{k}\}}}} \\
&= (t)^{(t/(t-1))\cdot(1-\ln(q)/\ln(t))},
\end{aligned}
$$

and

$$
\begin{aligned}
A_{\{\hat{k},\infty\}} &= (1 - c_{\{1,\hat{k}\}})^{\frac{(\hat{k}-1)}{\hat{k}}} \\
&= t^{-\left(1-\frac{\ln q}{\ln t}\right)}.
\end{aligned}
$$

Therefore,

$$A_{\{\hat{k},\infty\}}/A_{\{1,\hat{k}\}} = t^{-\left(1-\frac{\ln q}{\ln t}\right)\cdot\left(\frac{1}{1-t}\right)}.$$

This proves the Theorem.

Remark 7.1. We note that equation (25) gives the ratio of degree of uniformity of parts for general divisions at the $(1 - t)$-th percentile mark and can be used to derive specific empirical rules for specific percentages. This means that the last t% of the geometric distribution with indices $\{\hat{k}, \infty\}$ is $t^{-\left(1-\frac{\ln q}{\ln t}\right)\cdot\left(\frac{1}{1-t}\right)}$ as uniformly distributed as the first $(1 - t)$% with indices $\{1, \hat{k}\}$.

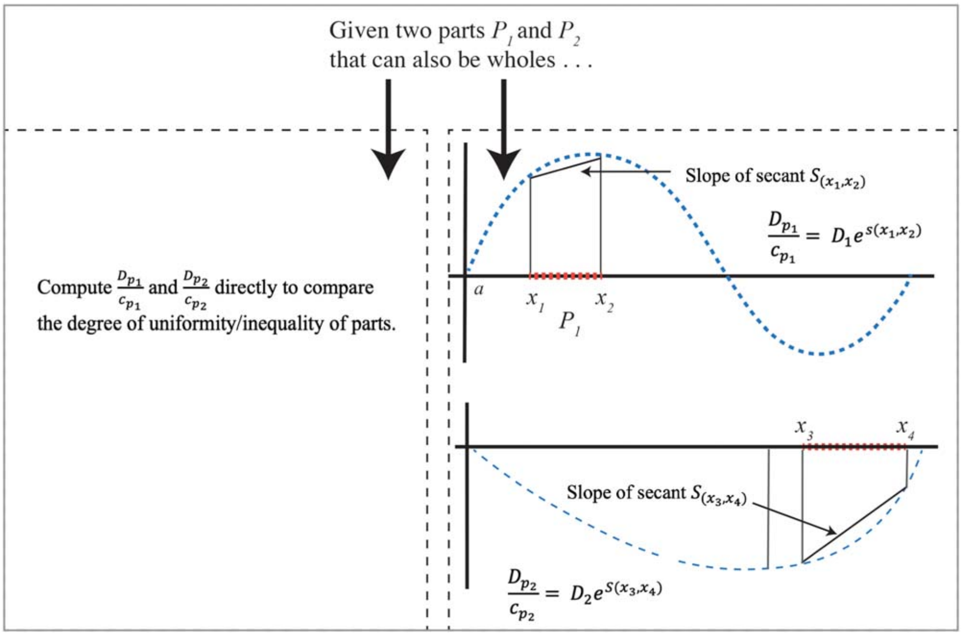

Figure 3. Picture showing the two ways of computing the degree of uniformity/inequality of parts of a distribution i.e. direct computation and from the slope of secants of the slope of diversity curve. The two graphs are the slope of diversity curves for parts P_1 and P_2 from two entirely different distributions.

We end by summarizing the general mechanics of quantification of the degree of uniformity/inequality of part or whole of one distribution with another. The heart of the mechanics comes from theorem 2.2. Assume we are given two parts $P_1 = (x_1, x_2)$ and $P_2 = (x_3, x_4)$. We have the following ways to quantify the degree of uniformity/inequality of said parts:

(i) We could directly compute the degree of uniformity/inequality of the parts by computing $\frac{D_{P_1}}{c_{P_1}}$ and $\frac{D_{P_2}}{c_{P_2}}$ from scratch. This is probably the best approach for distributions that don't admit analytical expressions for diversity, and also if we are only interested in the given parts P_1 and P_2

(ii) A more comprehensive approach would be to plot the *slope of diversity curve* for the entire distribution(s) and computing the slopes of secants of the parts $S_{(x_1,x_2)}$ and $S_{(x_3,x_4)}$. We could achieve this by simply drawing the secants and computing the respective slopes. This allows use to compute the degree of uniformity using $\frac{D_P}{c_P} = D \cdot e^{S_{(x_1,x_2)}}$, where D is the diversity of the entire distribution. This was shown in (Rajaram *et al* 2023, 2024). If the two parts P_1 and P_2 are both from the same distribution, then since D is the same for both parts (since they are both from the same distribution), we can form the ratio $\frac{A_{P_1}}{A_{P_2}}$, where $A_P = \frac{D_P}{c_P D}$ for parts P_1 and P_2. This is very easily computed using theorem 2.2 by $\frac{A_{P_1}}{A_{P_2}} = e^{S_{(x_1,x_2)} - S_{(x_3,x_4)}}$, since the total diversity D cancels out in this case. If the parts are from different distributions, then we can compute $\frac{\left(\frac{D_{P_1}}{c_{P_1}}\right)}{\left(\frac{D_{P_2}}{c_{P_2}}\right)} = \frac{D_1}{D_2} \cdot e^{S_{(x_1,x_2)} - S_{(x_3,x_4)}}$, where D_1 and D_2 are the diversities of the two different distributions respectively. The *slope of diversity* curve will now allow to compute the degree of uniformity/inequality of other parts of the distribution as well since all we need to do is compute the corresponding slopes of secants for those parts and repeat the above procedure.

We end by showing an illustration in figure 3 of the above explanation of the quantification of the degree of uniformity/inequality for the most general case.

8. Conclusions

We started the paper with three general questions.

1. Is there a universal way to measure inequality based purely on the entropic uncertainty that is inherent to probability distributions and nothing else? We have answered this question through our novel usage of case-

based entropy curve and in a more general sense, through the slope of diversity curve. The answer is that the degree of uniformity of parts P_i, as measured by the ratio $\frac{D_{P_i}}{c_{P_i}}$ from the slope of secant of the corresponding part in the slope of diversity curve, is a direct quantitative measure of how uniformly distributed a given part P_i is compared to another part P_j. Comparing this ratio for parts, as well as the whole if needed, is a direct way of measuring inequality from an entropic perspective, as this ratio measures the localization of inequality of the original distribution in the part P. We showed an application of this concept by looking at the slope of diversity curve and comparing the first half and the second half of the power law as an example in section 5.

2. Can we make it so this new method allows us to compare parts of a distribution and entire distributions alike? We answered this question through the usage of the slope of diversity curve to compare the parts of the power law distribution and the exponential distribution in section 6. The novelty of our approach is through the usage of the ratio of mathematical diversity for a given part P (denoted by D_P) and its cumulative probability (denoted by c_P), instead of the original random variable X_P, to compare equivalent parts in a scale free manner. Specifically, we were able to show that the first half of the exponential distribution is exactly half as uniform as the entire distribution and the second half is twice as uniform as the entire distribution. Also, that the second half of the exponential distribution is 4 times more uniformly distributed as the first half. For the power law, all these ratios are multiplied by the factor $\frac{\alpha}{\alpha-1}$ and hence lesser, in general compared to corresponding ratios for the exponential distribution. For general parts, the parts that are being compared need not have the same range of values for the cumulative probability c since we are more interested in computing the degree of inequality $\frac{D_P}{c_P}$ instead of just c_P.

3. Can we use this technique to glean out empirical rules similar to the Pareto principle, not just for the power law but for any given distribution or its parts thereof? We demonstrated this in section 4.1 where we analytically proved the so called $60-40$ rule that was observed for right-skewed distributions in (Rajaram *et al* 2023). This is just an example application to see which percentile of cases covers which entropic part of the distribution as measured by the case-based entropy. The entropic comparison using the slope of diversity curves for the power law and the exponential distribution also gleaned out empirical rules about ratio of degree of uniformity of parts of the respective distributions. Hence, in a more general sense, using the case-based entropy curve and the slope of diversity curve and looking at ratios of slopes of secants of parts will provide us with important and interesting comparisons of parts of a given distribution, as well as comparisons with other distributions leading to general empirical rules governing the distribution of degree of uniformity (or inequality) among said parts.

We chose the power law and the exponential distribution as analytical examples because they have been studied in depth in the literature and are also known to model several natural phenomena. We also used the geometric distribution as an example to demonstrate that the general machinery is versatile enough to handle discrete distributions. In the most general case, when dealing with empirical data, we surmise that explicit analytical expressions for the case-based entropy curve and the slope of diversity curve may not be feasible. However, given the computational formulas for these two curves as seen in the background material in section 2, the case-based entropy and the slope of diversity curves lend themselves to easy computation. In our future work, we will endeavor to create a tool that automatically computes these two curves from empirical data and also allows the user to draw secants, compute degree of uniformity of parts etc.

In this paper, we have used the exponential and power law distributions as examples to show how we can compare the distribution of degree of uniformity (or inequality) of their corresponding parts (as measured by the ratio $\frac{D_P}{c_P}$) using their corresponding slope of diversity curves. Apart from the power law and the exponential distribution, the rich plethora of distributions that one encounters in probability theory can now be viewed from the lens of mathematical diversity, using the case-based entropy and the slope of diversity curves. The demonstrations shown in this paper, using the aforementioned curves, lend themselves into important tools of analysis from the point of view of inequality and degree of uniformity of distributions and their parts. Analysis of distributions and their parts from the viewpoint of inequality or degree of uniformity is an important aspect of study of distributions. One application that comes to mind using the methods demonstrated in this paper is to identify and quantify parts of a given distribution (assuming the random variable being studied is a certain resource that is supposed to be distributed) based on their degree of uniformity relative the entire distribution. This will allow us to identify parts of the distribution which are plagued by inherent inequities that are not obviously discerned by simply looking at the shape of the distribution. In essence, we have devised a way to quantify inequality among parts of a distribution from an entropic perspective, and have also shown a way to compute the said quantification using the slope of diversity and the case-based entropy curves. Identification of parts of a distribution that are severely deprived of diversity in the form of degree of uniformity can lead to formulation of efficient policies that will result in distribution of resources that suits a certain diversity structure of parts.

In our future work, we will endeavor to construct a computational tool as mentioned above that will allow the user to input empirically obtained data and allow them to (a) plot the original probability distribution g_1, the case-based entropy curve g_2 and the slope of diversity curve g_3 with one-click, (b) compare the degree of uniformity of parts of the distribution by allowing the user to draw secant lines in the slope of diversity curve and look for other parts of the distribution that have the same or a scalar multiple of the slope of the secant line, and many more user-friendly routines that allow the user to analyze the empirical data from the lens of mathematical diversity and degree of uniformity. Such a computational tool will be invaluable to numerically compare and contrast the degree of inequality of parts of probability distributions that don't have explicit expressions for their densities, let alone for the slope of diversity curve. For example, in (Liang *et al* 2016) the probability density is the solution of a differential equation, and hence, in such situations, a computational tool will plot the slope of diversity curve directly, allowing the researcher to compare the degree of inequality of parts. We will also endeavor to develop a measure-theoretic framework to systematically partition the original probability distributions into parts that are ordered with increasing degree of uniformity (or inequality).

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

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

References

Castellani B and Rajaram R 2016 Complex systems and the limiting law of restricted diversity complex systems and the limiting law of restricted diversity *Complexity* **21** 99–112

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

Hill M 1973 Diversity and evenness: a unifying notation and its consequences Diversity and evenness: a unifying notation and its consequences *Ecology* **54** 27–432

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

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

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

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

Liang Y, Ye A Q, Chen W, Gatto R G, Colon-Perez L, Mareci T H and Magin R L 2016 A fractal derivative model for the characterization of anomalous diffusion in magnetic resonance imaging A fractal derivative model for the characterization of anomalous diffusion in magnetic resonance imaging *Commun. Nonlinear Sci. Numer. Simul.* **39** 529–37

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

Newman M 2005 Power laws, Pareto distributions and Zipf's law 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 '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 The measurement of species diversity *Annual Review of Ecological Systems* **5** 285–307

Rajaram R, Ritchey N and Castellani B 2023 Mathematical diversity of parts for a continuous distribution Mathematical diversity of parts for a continuous distribution *Journal of Physics Communications* **7** 075006

Rajaram R, Ritchey N and Castellani B 2024 On the comparison of diversity of parts of a distribution On the comparison of diversity of parts of a distribution *Journal of Physics Communications* **8** 025008