The Mathematics of Diversity

TOM LEINSTER

The challenge of pinning down the concept of diversity leads to mathematics of surprising depth and breadth, touching areas from geometric measure to information theory to functional equations. It also fits into a wide-ranging category-theoretic story that unifies invariants of size across mathematics.

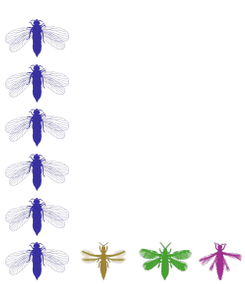
We often hear that the decline in biodiversity is one of the world's most urgent crises, or talk about diversity in the workplace, but what does 'diversity' actually mean? If someone says that the diversity of British birds has dropped by 20%, what is the quantity being measured?

To a pure mathematician, these questions might seem of peripheral interest. However, pursuing them leads to mathematical riches. For example, the theory of diversity produces a canonical probability measure on any compact metric space, unifying Haar and Lebesgue measures. There are also connections with classical geometric quantities such as volume, surface area and dimension, and with information entropy. And we will see that diversity is in some sense the probabilistic analogue of cardinality, ultimately glimpsing a bigger category-theoretic picture that links up size-like invariants across mathematics.

What is diversity?

The difficulties of quantifying diversity are not just practical or statistical; it is a *conceptual* challenge too. Even assuming perfect information, it is not clear what it should mean for one community to be more diverse than another.

Consider these two insect communities:



Community A

Community B

Community A has four species, but one is dominant and the other three rare. Community B has only three species, but they are equally abundant. Is community A more diverse, because it has more species? Or is it community B, because it is evenly balanced?

Ecologists have debated questions such as these since the 1940s. There have been parallel debates in genetics and economics. Because 'diversity' has many possible meanings, many dozens of diversity measures have been proposed.

For a mathematician, the first challenge is to bring some order to the bewildering array of proposed definitions of diversity. For example, here are four of the most commonly used; the reader is invited to spot the pattern.

In all of them, we start from an ecological community consisting of n species, writing p_i for the relative abundance of the ith species. 'Relative' means $\sum p_i = 1$, so that $\mathbf{p} = (p_1, \ldots, p_n)$ is a finite probability distribution.

The simplest of all diversity measures is the number of species present. This is the cardinality

$$D_0(\mathbf{p}) = \big|\{i : p_i > 0\}\big|.$$

Also popular is

$$D_1(\mathbf{p}) = 1/p_1^{p_1} p_2^{p_2} \cdots p_n^{p_n}$$

(where $0^0 = 1$). This is the exponential of the Shannon entropy, $-\sum p_i \log p_i$, central to information theory. Another frequently used diversity measure is

$$D_2(\mathbf{p}) = 1\Big/ \sum_i p_i^2,$$

the reciprocal of the probability that two random individuals are of the same species. Finally,

$$D_\infty(\mathbf{p}) = 1/\max_i p_i$$

reflects the extent to which the community is dominated by a single species. For example, it is not much more than 1 in Community A above.

As the notation suggests, these four diversity measures are all members of a single family $(D_q)_{0 \leq q \leq \infty}$. The general formula is

$$D_q(\mathbf{p}) = \left(\sum_{i:\, p_i > 0} p_i^q \right)^{1/(1-q)}. \tag{1}$$

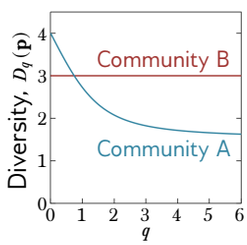
At $q = 1$, this formula is interpreted as $\lim_{q \to 1} D_q(\mathbf{p})$, which turns out to be $D_1(\mathbf{p})$ as defined above. The same goes for $q = \infty$.

Ecologists call D_q the **Hill number** of order q, and to information theorists it is the exponential of the **Rényi entropy** of order q.

What does it mean for a community to have diversity 16.2, say? To answer this, we use a key property of the Hill numbers: that $D_q(1/n, \ldots, 1/n) = n$ for all n and q. So, $D_q(\mathbf{p}) = 16.2$ means our community is slightly more diverse than a community of 16 equally common species.

Different values of the parameter q embody different viewpoints on what diversity is, and make different judgements on when one community is more diverse than another. For example, the Hill number of order 0 judges Community A to be more diverse than Community B, because it has more species, but the judgement is reversed at $q = \infty$.

To avoid cherry-picking, it is best to calculate the Hill numbers of *all* orders. The graph of $D_q(\mathbf{p})$ against q is called the **diversity profile** of $\mathbf{p}$, and is shown here for our two insect communities.



The axiomatic approach

What makes the Hill numbers special? Why *this* formula? The most compelling answer is a theorem: they are the only measures that behave logically. In other words, any other diversity measure will contravene at least one of a list of properties embodying common intuitions about what diversity means.

Some of these properties are mundane. For instance, $D_q(\mathbf{p})$ is unchanged if we permute the order of the species or add a new species with zero abundance. The diversity of a community is between 1 and the number of species. And although it is not quite true that the Hill numbers are all continuous in $\mathbf{p}$ (the exception being $D_0(\mathbf{p})$), they are continuous when restricted to distributions $\mathbf{p}$ with $p_i > 0$ for all i.

The last two properties take a little more explanation. Suppose our community is spread across k islands, and that no species is present on more than one island. One property is that if we increase the diversity on one island and leave the rest the same, then the diversity of the whole does not decrease. The other is that if the islands have the same total populations and the same number of species distributed in the same way, then the diversity of the whole community is k times the diversity of each island.

The theorem is as follows [4, Section 7.4]. Consider a hypothetical diversity measure D assigning a positive real number $D(\mathbf{p})$ to each finite probability distribution $\mathbf{p}$. Then D satisfies all of the properties above if and only if D is the Hill number D_q for some $q \in [0, \infty]$.

The moral is that diversity measures can be treated *axiomatically*. Before one settles on a definition of 'diversity', one should decide what *properties* one believes a diversity measure should have. And the decisions made above force us to use a particular family of measures: the Hill numbers.

The fact that the Hill numbers unify some of the most commonly used diversity measures in ecology demonstrates that the axiomatic approach is not some mathematical nicety, but connects directly to natural science.

Similarity and metrics

The problem with the Hill numbers is that they take no account of the varying similarities between species. Intuitively, we would regard a field of ten similar species of grass as less diverse than if it had ten very different species. We therefore refine our model.

As before, our community is divided into n species with relative abundances $\mathbf{p} = (p_1, \ldots, p_n)$, but now we also take as given a similarity coefficient $Z_{ij} \geq 0$

between each pair (i,j) of species, forming a matrix $Z = (Z_{ij})_{1\leq i,j\leq n}$. A similarity of 0 means that species i and j have nothing in common, and a similarity of 1 that they are identical.

There are many ways to quantify similarity. The most naive is to take Z to be the identity matrix I: different species are always completely dissimilar. Less crudely, Z_{ij} may measure percentage genetic similarity, or be determined phylogenetically or taxonomically. But for mathematics, a particularly significant choice—as we will see—is to start with a metric d on $\{1,\dots,n\}$ and put $Z_{ij} = e^{-d(i,j)}$.

The Hill numbers can now be upgraded to take species similarity into account. For an individual of species i, its expected similarity to an individual chosen at random is

$$(Z\mathbf{p})_i := \sum_{j=1}^{n} Z_{ij}p_j.$$

Then $(Z\mathbf{p})_i$ measures how ordinary this individual is within the community, so $1/(Z\mathbf{p})_i$ reflects its distinctiveness. To measure the diversity of the whole community, we take the average distinctiveness over all individuals—specifically, the power mean of order $1-q$, where again $0 \leq q \leq \infty$:

$$D_q^Z(\mathbf{p}) = \left(\sum_{i\colon p_i>0} p_i (Z\mathbf{p})_i^{q-1} \right)^{1/(1-q)}.$$

For example, the diversity of order 2 is

$$D_2^Z(\mathbf{p}) = 1\Big/ \sum_{i,j} p_i Z_{ij} p_j,$$

the reciprocal of the expected similarity between a random pair of individuals. As before, the cases $q = 1,\infty$ are interpreted by taking limits.

The *formula* for the diversity measures D_q^Z is less important than their *properties*. For instance, they reduce to the Hill numbers when $Z = I$, and for general Z they satisfy similarity-sensitive versions of the good properties of the Hill numbers. They are also highly versatile: if Z_{ij} measures genetic similarity then D_q^Z measures genetic diversity, if Z_{ij} is phylogenetic similarity then D_q^Z measures phylogenetic diversity (which may be different), etc. And these measures have been used beyond ecology in the humanities, social sciences, and computer science. But let us instead turn back towards the mathematics.

Maximizing diversity: three miracles

How diverse can a community be? If someone gives you a list of species and allows you to choose their relative abundances, how should you do so in order to maximize diversity?

If the varying similarities between species are ignored then the answer is simple: the uniform distribution maximizes diversity. In fact, for all $q \in [0,\infty]$, the Hill number D_q takes its maximum value of n at the uniform distribution $(1/n,\dots,1/n)$. But when similarity is taken into account, the plot thickens.

To see this intuitively, consider a hypothetical community consisting of two similar species of toad and one species of newt (Figure 1). The uniform distribution would make the community $2/3$ toad and $1/3$ newt, which does not seem maximally diverse. We would expect the maximal diversity to be achieved by a distribution closer to $(1/4,1/4,1/2)$, with the exact proportions depending on the degrees of similarity.

Figure 1. Three-species community. Placement indicates species similarity

Generally, given a similarity matrix Z representing a fixed list of species, which probability distributions $\mathbf{p}$ maximize the diversity $D_q^Z(\mathbf{p})$?

Since different values of the parameter q make different judgements on when one community is more diverse than another, we would expect the answer to vary with q. One distribution might maximize diversity of order 1, while another maximizes diversity of order 2.5, and so on.

But here a miracle occurs. There is a best of all possible worlds: a probability distribution $\mathbf{p}$ that maximizes $D_q^Z(\mathbf{p})$ for all $q \in [0,\infty]$ simultaneously [6]. (Typically, although not invariably, there is only one.) A second miracle: if $\mathbf{p}$ maximizes $D_q^Z(\mathbf{p})$ for *some* $q > 0$ then it maximizes $D_q^Z(\mathbf{p})$ for *all* $q \geq 0$. And a third: the *value* of the maximum diversity, $\sup_{\mathbf{p}} D_q^Z(\mathbf{p})$, is the same for all q.

These results make only minimal assumptions on the similarity matrix Z: it can be any symmetric matrix of nonnegative reals, nonzero on the diagonal.

An example: if our two toad species are 90% similar to each other and 40% similar to the newt then the distribution $\mathbf{p} = (0.261, 0.261, 0.478)$ maximizes $D_q^Z(\mathbf{p})$ for all q, in line with our expectations.

Or, take a graph G with vertices labelled $1, \dots, n$, define Z_{ij} to be 1 if $i = j$ or there is an edge between i and j, and 0 otherwise. Then diversity is maximized by the uniform distribution on any independent set of vertices of maximal size, where 'independent' means that no two vertices in the set have an edge between them. The value of the maximum diversity is the cardinality of a maximal independent set.

There is an algorithm that computes the maximum diversity distributions of any similarity matrix. However, the graph example implies that no algorithm can run in polynomial time in n, at least assuming **P** ≠ **NP**, because it is known that the same is true for the problem of determining a maximal independent set in a graph.

Geometry and measure theory

Finite spaces are all very well, but when we step up to compact spaces, we see that diversity is intimately related to geometry and measure.

The general context is a compact Hausdorff space X with a suitable kernel $Z\colon X \times X \to \mathbb{R}$, but for simplicity, let us assume that X is a metric space and $Z(x, y) = e^{-d(x,y)}$. The idea is that the 'diversity' of a probability measure μ on X should quantify how spread out μ is across X (Figure 2). In the finite case, $\mu = \mathbf{p}$ is a probability measure on $X = \{1, \dots, n\}$.

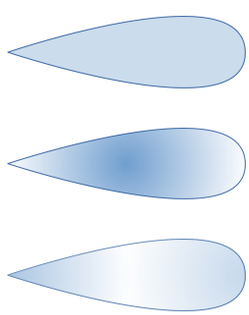


Figure 2. Three probability measures on the same space. Shading indicates concentration of measure

Formally, given a probability measure μ on X, the function

$$Z\mu\colon x \mapsto \int_X e^{-d(x,y)}\, d\mu(y)$$

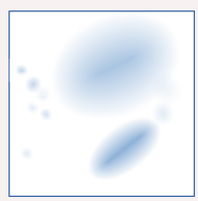
describes how typical each point x is; it is highest where the measure is concentrated the most. Hence $1/(Z\mu)(x)$ describes atypicality. The **diversity of order** q of μ is the average atypicality over the whole space, again taking 'average' in the sense of a power mean: for $q \in [0, \infty]$, we put

$$D_q^X(\mu) = \left(\int_X (1/Z\mu)^{1-q}\, d\mu \right)^{1/(1-q)},$$

taking limits as usual to make sense of this at $q = 1, \infty$. This generalizes the definition for finite spaces.

Spaces of microbes

It might be thought that infinite spaces of species would have no relevance in ecology. However, in microbial ecology, there is often no useful species classification and it is preferable to view the space of possible species as a continuum.



For instance, new flu vaccines are needed every year because the virus evolves rapidly, its genome constantly moving across the space of all conceivable genomes. Annual vaccines are designed to cover the major clusters (serotypes or strains) of that year's virus. The more diverse the virus is that year, the harder the vaccine-builders' job becomes. The figure shows a simplified, hypothetical virus community; the square is the space of all possible genetic types, with darker areas indicating the more abundant types.

The three miracles of maximum diversity on finite spaces extend verbatim to the compact setting [9, Section 7]. So, every compact metric space X carries a **maximizing measure**: a probability measure that maximizes D_q^X for all $q \in [0, \infty]$ at once. Also, the maximum diversity is the same for all q. This maximum diversity $D_{\max}(X) = \sup_\mu D_q^X(\mu) \in \mathbb{R}^+$, defined independently of q, is a real invariant of compact metric spaces. It expresses how much room a measure has to spread out in X.

Example 1. On a line segment $[0, \ell]$, the unique probability measure that maximizes diversity is the normalization of

$$\delta_0 + \lambda_{[0,\ell]} + \delta_\ell,$$

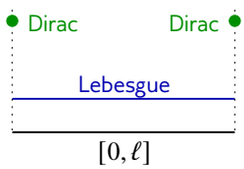
where δ_0 and δ_ℓ are the Dirac deltas at the endpoints and $\lambda_{[0,\ell]}$ is the restriction of Lebesgue measure.

The maximum diversity is $D_{\max}(X) = 1 + \frac{1}{2}\ell$. Each of the endpoints contributes $1/2$ to the maximum diversity, and the contribution of the interior is proportional to its length.

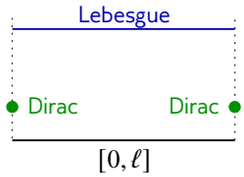
Beyond this example, there are very few spaces whose maximum diversity is known. It is unknown even for Euclidean balls of dimension greater than 1. However, the asymptotic behaviour of maximum diversity—how it behaves as distances are scaled to ∞—is somewhat understood and extremely interesting.

For a start, maximum diversity determines fractal dimension. Write tX for the metric space X with distances scaled up by a factor of $t > 0$. As Meckes showed, the growth rate of the function $t \mapsto D_{\max}(tX)$ is equal to the Minkowski dimension of X [10, Theorem 7.1]. So, for instance, if $D_{\max}(tX)$ grows like t^2 as $t \to \infty$ then X must be 2-dimensional.

A generic compact metric space X has a unique maximizing measure. In some sense, this is the canonical probability measure on X. However, it is scale-dependent. For example, the maximizing measure on a line segment gives more weight to the endpoints when the line is short than when it is long (Figure 3).



(a) Small ℓ



(b) Large ℓ

Figure 3. Figurative illustration of the maximum diversity measure on a line of length ℓ

But we can obtain a canonical *scale-free* measure on a space X by scaling to infinity, as shown in work with Roff [9]. Assuming that tX has a unique maximizing measure μ_{tX} for all $t \gg 0$, the **uniform measure** on X is $\lim_{t\to\infty} \mu_{tX}$ (if it exists).

Example 2. For a finite metric space X, the uniform measure gives probability $1/|X|$ to each point.

Example 3. When X is a compact subset of $\mathbb{R}^n$, the uniform measure is Lebesgue measure λ_n restricted to X and normalized to a probability measure (assuming that $\lambda_n(X) \neq 0$). What is remarkable about this result [9, Theorem 9.9] is that the maximizing measures μ_{tX} at *finite* scales t are all unknown, but $\lim_{t\to\infty} \mu_{tX}$ can still be found.

Example 4. When the isometry group of X acts transitively on points, the uniform measure on X is the Haar probability measure [9, Proposition 9.5].

Thus, the concept of uniform measure that emerges from maximum diversity provides a unification of Lebesgue measure, Haar measure, and uniform measure on finite sets.

Much about uniform measure remains unknown. For instance, examples suggest that whenever $A \subseteq X$ is the closure of its interior, the uniform measure of A depends only on the isometry types of A and X, not how A is embedded in X. Specifically, it appears that the uniform measure on X gives A a probability of

$$\lim_{t\to\infty} D_{\max}(tA)/D_{\max}(tX).$$

But so far these are only empirical observations, not theorems.

Zooming out: the magnitude programme

The story told here is a fragment of a much larger one. The *magnitude programme* [8] is a sweeping category-theoretic study of notions of size in mathematics, bringing together geometric, algebraic, topological, measure-theoretic and graph-theoretic invariants, as well as diversity.

The fundamental concept of this programme is the *magnitude of an enriched category*. Enriched categories are highly general structures, including both ordinary categories and metric spaces. The magnitude of an ordinary category is closely related to the topological concept of Euler characteristic, and the magnitude of a metric space is a real invariant of size that had barely been explored before.

The magnitude $|X|$ of a metric space X is closely related to its maximum diversity. Loosely, it is what maximum diversity would be if negative probabilities or species abundances were allowed. But magnitude is also closely related to classical geometric measures. Using increasingly sophisticated techniques, analysts have proved that the function $t \mapsto |tX|$ determines the dimension, volume, surface area, perimeter, etc., of a space X [8, 1, 10]. In fact, the notion of maximum diversity was an essential part of some of these

proofs, establishing the worth of diversity as a pure-mathematical tool.

Much more on the magnitude programme can be found in [7] and [8], including its accompanying homology theory [2], a rapidly blossoming collection of applications, and an ever-growing network of unexpected connections both within and beyond mathematics. The new world of magnitude and diversity is full of open questions. It contains whole continents on which we have hardly set foot, waiting to be explored.

FURTHER READING

[1] H. Gimperlein, M. Goffeng, On the magnitude function of domains in Euclidean space, American Journal of Mathematics 143 (2021), 939–967.
[2] R. Hepworth, S. Willerton, Categorifying the magnitude of a graph, Homology, Homotopy and Applications 19 (2017), 31–60.
[3] T. Leinster, The magnitude of metric spaces, Doc. Math. 18 (2013), 857–905.
[4] T. Leinster, Entropy and Diversity: The Axiomatic Approach, Cambridge University Press, 2021.
[5] T. Leinster and C. Cobbold, Measuring diversity: the importance of species similarity, Ecology 93 (2012), 477–489.
[6] T. Leinster, M. Meckes, Maximizing diversity in biology and beyond, Entropy 18 (2016), 88.
[7] T. Leinster, M. Meckes, The magnitude of a metric space: from category theory to geometric measure theory, in: N. Gigli (ed.), Measure theory in Non-Smooth Spaces, De Gruyter Open, 2017, 156–193.
[8] T. Leinster, M. Meckes, Magnitude: a bibliography, maths.ed.ac.uk/~tl/magbib, 2018–2026.
[9] T. Leinster, E. Roff, The maximum entropy of a metric space, Q. J. Math 72 (2021), 1271–1309.
[10] M. Meckes, Magnitude, diversity, capacities, and dimensions of metric spaces, Potential Anal. 42 (2015), 549–572.

Tom Leinster

Tom is a mathematician at the University of Edinburgh. He is interested in category theory and its more unusual applications.