The maximum entropy of a metric space

Tom Leinster∗  Emily Roff†

We define a one-parameter family of entropies, each assigning a real number to any probability measure on a compact metric space (or, more generally, a compact Hausdorff space with a notion of similarity between points). These entropies generalise the Shannon and Rényi entropies of information theory.

We prove that on any space \( X \), there is a single probability measure maximising all these entropies simultaneously. Moreover, all the entropies have the same maximum value: the maximum entropy of \( X \). As \( X \) is scaled up, the maximum entropy grows; its asymptotics determine geometric information about \( X \), including the volume and dimension. We also study the large-scale limit of the maximising measure itself, arguing that it should be regarded as the canonical or uniform measure on \( X \).

Primarily we work not with entropy itself but its exponential, called diversity and (for finite spaces) used as a measure of biodiversity. Our main theorem was first proved in the finite case by Leinster and Meckes [14].

Contents

1 Introduction 1
2 Topological and analytic preliminaries 6
3 Typicality 11
4 Diversity and entropy 14
5 Magnitude 21
6 Balanced and maximising measures 24
7 The maximisation theorem 28
8 Metric spaces 31
9 The uniform measure 34
10 Open questions 38

1 Introduction

This paper introduces and explores a largely new invariant of compact metric spaces, the maximum entropy. Intuitively, it measures how much room a probability distribution on the space has available to spread out.

∗School of Mathematics, University of Edinburgh; Tom.Leinster@ed.ac.uk
†School of Mathematics, University of Edinburgh; Emily.Roff@ed.ac.uk
Maximum entropy has several claims to importance. First, it is the maximal value of not just one measure of entropy, but an uncountable infinity of them. It is a theorem, proved here, that they all have the same maximum. Second, these entropies have already been found useful and meaningful in the life sciences, where they (or rather their exponentials) are interpreted as measures of biological diversity [13, 26].

Third, the exponential of maximum entropy—called maximum diversity—plays a similar role for metric spaces as cardinality does for sets. Like cardinality, it increases when the space is made bigger (either by adding new points or increasing distances), and in the special case of a finite space where all distances are $\infty$, it is literally the cardinality. Maximum diversity at large scales is also closely related to volume and dimension (themselves analogues of cardinality), as explained below.

**Measuring diversity** The backdrop for the theory is a compact Hausdorff topological space $X$, equipped with a way to measure the similarity between each pair of points. This data is encoded as a similarity kernel: a continuous function $K : X \times X \to [0, \infty)$ taking strictly positive values on the diagonal. We call the pair $(X, K)$ a space with similarities.

In a metric space, we naturally view points as similar if they are close together, and we define a similarity kernel $K$ by $K(x, y) = e^{-d(x, y)}$. Of course, there are other possible choices of kernel, but this particular choice proves to be a wise one for reasons explained in Example 3.3.

For simplicity, in this introduction we focus on the case of metric spaces rather than fully general spaces with similarity.

We would like to quantify the extent to which a probability distribution on a metric space is spread out across the space, in a way that is sensitive to distance. A thinly spread distribution will be said to have ‘high diversity’, or equivalently ‘high entropy’. Definitions are given later; here we just describe the intuitive idea.

Figure 1 depicts three distributions on the same space. Distribution (a) is the least diverse, with most of its mass concentrated in a small region. Distribution (b) is uniform, and might therefore seem to be the most diverse or spread out distribution possible. However, there is an argument that distribution (c) is
more diverse. In moving from (b) to (c), some of the mass has been pushed out to the ends, so a pair of points chosen at random according to distribution (c) may be more likely to be far apart than when chosen according to (b).

One can indeed define diversity in terms of the expected proximity between a random pair of points. But that is just one of an infinite family of ways to quantify diversity, each of which captures something slightly different about how a distribution is spread across the space.

To define that family of diversity measures, we first introduce the notion of the typicality of a point with respect to a distribution. Given a compact metric space $X$, a probability measure $\mu$ on $X$, and a point $x \in X$, we regard $x$ as 'typical' of $\mu$ if a point chosen at random according to $\mu$ is usually near to $x$, and 'atypical' if it is likely to be far away. Formally, define a function $K\mu$ on $X$ by

$$(K\mu)(x) = \int e^{-d(x, \cdot)} \, d\mu.$$  

We call $(K\mu)(x)$ the typicality of $x$, and $1/(K\mu)(x)$ its atypicality.

A distribution is widely spread across $X$ if most points are distant from most of the mass—that is, if the atypicality function $1/K\mu$ takes large values on most of $X$. A reasonable way to quantify the diversity of a probability measure $\mu$, then, is as the average atypicality of points in $X$.

Explicitly, for $q \neq 1, \pm \infty$, we define the diversity of order $q$ of $\mu$ to be

$$D^K_q(\mu) = \left( \int (1/K\mu)^{1-q} \, d\mu \right)^{1/(1-q)},$$

while at $q = 1$ and $q = \pm \infty$ this expression takes its limiting values. The entropy $H^K_q(\mu)$ of order $q$ is $\log D^K_q(\mu)$: entropy is the logarithm of diversity.

**Diversity and entropy** Any finite set can be given the structure of a compact metric space by taking all distances between distinct points to be $\infty$. The similarity kernel $K = e^{-d(\cdot, \cdot)}$ is then the Kronecker delta $\delta$. In this trivial case, the entropy $H^K_1$ is precisely the Rényi entropy of order $q$, well-known in information theory. In particular, $H^K_1$ is Shannon entropy.

Entropy is an important quantitative and conceptual tool in many fields, including in mathematical ecology, where the exponentials $D^K_q$ of the Rényi entropies are known as the Hill numbers and used as measures of biological diversity [8]. However, the Hill numbers have a serious deficiency. They fail to reflect a fundamental intuition about diversity, namely that, all else being equal, a biological community is regarded as more diverse when the species are very different than when they are very similar.

To repair this deficiency, one can equip the set of species in an ecological community with a kernel (or matrix) $K$ recording their pairwise similarities. The choice $K = \delta$ represents the crude assumption that each species in the community is completely dissimilar to each other species. Using this data, one
can define generalised Hill numbers, sensitive to species similarity. These are the similarity-sensitive diversity measures of [13], of which our measures are the continuous generalisation.

A maximisation theorem Crucially, when comparing the diversity of distributions, different values of the parameter $q$ lead to different results. That is, given a collection $M$ of probability distributions on a metric space and given distinct $q, q' \in [0, \infty]$, the diversities $D^K_q$ and $D^K_{q'}$ generally give different orderings to the elements of $M$. Examples in the biological setting can be found in Section 5 of [13].

The surprise of our main theorem (Theorem 7.1) is that when it comes to maximising diversity on a compact metric space $X$, there is consensus: there is guaranteed to exist some probability measure $\mu$ on $X$ that maximises $D^K_q(\mu)$ for every nonnegative $q$ at once. Moreover, the diversity of order $q$ of a maximising distribution is the same for all $q \in [0, \infty]$. Thus, one can speak unambiguously about the maximum diversity of a compact metric space $X$—defined to be

$$D_{\text{max}}(X) = \sup_{\mu} D^K_q(\mu)$$

for any $q \in [0, \infty]$—knowing that there exists a probability distribution attaining this supremum for all orders $q$.

Theorem 7.1 extends to compact spaces a result that was established for finite spaces in [14]. (Note that the maximising measure on a finite metric space is not usually uniform.) While the proof of the result for compact spaces follows broadly the same strategy as in the finite case, substantial analytic issues arise.

Geometric connections The maximum diversity theorem has geometric significance, linking diversity measures to the intrinsic volumes of classical convex geometry and to geometric measure theory. Roughly speaking, maximum diversity provides a measure of the size of a metric space.

More specifically, Corollary 7.4 of our main theorem connects maximum diversity with another, more extensively studied invariant of a metric space: its magnitude. First introduced as a generalised Euler characteristic for enriched categories ([11, 12]), magnitude specialises to metric spaces by way of Lawvere’s observation that metric spaces are enriched categories [10]. The magnitude $|X| \in \mathbb{R}$ of a metric space $X$ captures a rich variety of classical geometric data, including some intrinsic volumes of Riemannian manifolds and of compact sets in $\ell^1_1$ and $\mathbb{R}^n$. The definition of magnitude and a few of its basic properties are given in Sections 5 and 8 below; a detailed survey can be found in [15].

In Sections 6 and 7 we show that the maximum diversity of a compact space is the magnitude of a certain subset: the support of any maximising measure. In Section 8 we use this fact, and known facts about magnitude, to establish a handful of examples of maximum diversity for metric spaces.

Many results on magnitude are asymptotic, in the following sense. Given a space $X$ with metric $d$, and a positive real number $t$, define the scaled metric

$$d^t(x, y) = \frac{d(x, y)}{t}$$

for $t > 0$. As $t \to \infty$, the magnitude $|X|_t$ of the scaled space $X$ goes to zero, and $|X|_t$ is asymptotically small for $t$ large. The precise dependence of $|X|_t$ on $t$ is the content of a number of known results.
space \( tX \) to be the set \( X \) equipped with the metric \( t \cdot d \). It has proved fruitful to consider, for a fixed metric space \( X \), the entire family of spaces \( (tX)_{t>0} \) and the (partially-defined) magnitude function \( t \mapsto |tX| \). For instance, in [1], Barceló and Carbery showed that the volume of a compact subset of \( \mathbb{R}^n \) can be recovered as the leading term in an asymptotic expansion of its magnitude function, while in [5], Gimperlein and Goffeng showed (subject to technical conditions) that lower order terms capture surface area and the integral of mean curvature.

Given this, and given the relationship between magnitude and maximum diversity, it is natural to consider the function \( t \mapsto D_{\text{max}}(tX) \). Indeed, the asymptotic properties of maximum diversity have already been shown to be of geometric interest. In [17], Meckes defined the maximum diversity of a compact metric space to be the maximum value of its diversity of order 2, and used this definition—now vindicated by our main theorem—to prove the following relationship between maximum diversity and Minkowski dimension:

**Theorem 1.1** (Meckes, [17], Theorem 7.1) Let \( X \) be a compact metric space, and let \( \dim_{\text{Mink}}(X) \) denote the Minkowski dimension of \( X \). Then

\[
\lim_{t \to \infty} \frac{\log D_{\text{max}}(tX)}{\log t} = \dim_{\text{Mink}}(X),
\]

and the left-hand side is defined if and only if the right-hand side is defined.

That is, the Minkowski dimension of \( X \) is the growth rate of \( D_{\text{max}}(tX) \) for large \( t \). Theorem 9.7 below is a companion result for the volume of sets \( X \subseteq \mathbb{R}^n \):

\[
\lim_{t \to \infty} \frac{D_{\text{max}}(tX)}{t^n} \propto \text{Vol}(X).
\]

In short, maximum diversity determines dimension and volume.

**Entropy and uniform measure** Taking logarithms throughout, the maximum diversity theorem tells us that every compact metric space admits a probability measure maximising the entropies \( H^K_q \) of all orders \( q \) simultaneously. Statisticians have long recognised that maximum entropy distributions are special. However, it is not necessarily helpful to view the maximum entropy measure on a metric space \( X \) as being the ‘canonical measure’, since it is not scale-invariant: if we multiply all distances in \( X \) by a constant factor \( t \), the maximising measure changes.

In Section 9 we propose a canonical, scale-invariant, choice of probability measure on each of a wide class of metric spaces, and call it the uniform measure. It is the limit as \( t \to \infty \) of the maximum entropy measure on \( tX \). We use the examples of maximum diversity established in Section 8 to show that, in several familiar cases, this definition succeeds in capturing the classical notion of uniform distribution.

**Conventions** Throughout, a measure on a topological space means a Radon measure. All measures are positive, unless stated otherwise. A function \( f : \)
\( R \to R \) is **increasing** if \( f(y) \leq f(x) \) for all \( y \leq x \), and **strictly increasing** if \( f(y) < f(x) \) for all \( y < x \); **decreasing** and **strictly decreasing** are used similarly.

**Acknowledgements**  We thank Mark Meckes for many useful conversations, and especially for allowing us to include Proposition 8.8, which is due to him. Thanks also to Christina Cobbold for helpful suggestions.

2 Topological and analytic preliminaries

In this paper we are concerned with properties of probability measures on a topological space. We begin by collecting the key topological and measure-theoretic facts that will be needed, also taking the opportunity to fix notation.

**Topologising spaces of functions**  Let \( X \) and \( Y \) be topological spaces. We write \( \text{Top}(X,Y) \) for the set of continuous functions from \( X \) to \( Y \), which can be topologised as follows. For subsets \( K \subseteq X \) and \( U \subseteq Y \), write

\[
F(K,U) = \{ f \in \text{Top}(X,Y) : fK \subseteq U \}.
\]

The **compact-open topology** on \( \text{Top}(X,Y) \) is the topology generated by \( F(K,U) \) for all compact \( K \subseteq X \) and open \( U \subseteq Y \); its properties are described in Section 46 of [19].

The most important property of the compact-open topology involves **locally compact** spaces, that is, those in which every neighbourhood of a point contains a compact neighbourhood. Every compact Hausdorff space is locally compact.

**Proposition 2.1**  Let \( Y \) be a locally compact space, and let \( X \) and \( Z \) be any topological spaces. A map \( f : X \times Y \to Z \) is continuous if and only if the map \( \overline{f} : X \to \text{Top}(Y,Z) \) given by \( \overline{f}(x)(y) = f(x,y) \) is continuous with respect to the compact-open topology.

**Proof.**  This is Proposition 7.1.5 in [2].

Categorically, this states that locally compact spaces \( Y \) are exponentiable: the functor \(- \times Y : \text{Top} \to \text{Top}\) has a right adjoint, given by \( \text{Top}(Y,-) \) with the compact-open topology.

Now let \( X \) be a topological space and \( Y \) a metric space. The set \( \text{Top}(X,Y) \) carries the metric \( d_{\infty} \) given by

\[
d_{\infty}(f,g) = \sup_{x \in X} d(f(x),g(x)).
\]

The **uniform topology** on \( \text{Top}(X,Y) \) is the topology induced by this metric.

**Lemma 2.2**  Let \( X \) be a compact topological space and \( Y \) a metric space. Then the compact-open and uniform topologies on \( \text{Top}(X,Y) \) are equal.
Proof. This follows from Theorems 46.7 and 46.8 in [19].

We will only use function spaces \( \text{Top}(X,Y) \) in which \( X \) is compact Hausdorff and \( Y \) is metric, and we always understand \( \text{Top}(X,Y) \) to come equipped with the unambiguous topology of Lemma 2.2. The case \( Y = \mathbb{R} \) is especially important, and we write \( \text{Top}(X) = \text{Top}(X, \mathbb{R}) \). The metric \( d_\infty \) on \( C(X) \) is induced by the uniform norm \( \| \cdot \|_\infty \) on \( C(X) \), defined by \( \| f \|_\infty = \sup_{x \in X} |f(x)| \).

The uniform topology on function spaces is functorial in the following sense.

**Lemma 2.3** Let \( X \) be a compact topological space, \( Y \) and \( Y' \) metric spaces, and \( \phi : Y \rightarrow Y' \) a continuous function. Then the induced map \( \phi \circ - : \text{Top}(X,Y) \rightarrow \text{Top}(X,Y') \) is continuous with respect to the uniform topology on the domain and codomain.

**Proof.** It is elementary consequence of the definitions that \( \phi \circ - \) is continuous with respect to the compact-open topology on the domain and codomain. The result then follows from Lemma 2.2.

The preceding results imply:

**Corollary 2.4** Let \( X \) be any topological space, \( Y \) a compact Hausdorff space, and \( Z \) a metric space. A map \( f : X \times Y \rightarrow Z \) is continuous if and only if the map \( \bar{f} : X \rightarrow \text{Top}(Y,Z) \) is continuous with respect to the uniform topology.

### Topologising spaces of measures

**Until further notice, let X denote a compact Hausdorff space.**

Equip the vector space \( C(X) \) with the uniform norm \( \| \cdot \|_\infty \). By the Riesz representation theorem, its topological dual \( C(X)^* \) is isomorphic to the space \( M(X) \) of signed measures on \( X \). The dual norm on \( M(X) \) is the total variation norm, \( \| \mu \| = |\mu|(X) \), and the pairing corresponding to the isomorphism is \( \langle f, \mu \rangle = \int_X f \, d\mu \) (where \( f \in C(X), \mu \in M(X) \)).

Being a dual vector space, \( M(X) \) carries the weak* topology, which is the coarsest topology such that the map \( \langle f, - \rangle : M(X) \rightarrow \mathbb{R} \) is continuous for each \( f \in C(X) \). Whenever we invoke a topology on \( M(X) \) or one of its subsets, we will always mean this one. It is Hausdorff, and the Banach–Alaoglu theorem states that closed bounded subsets of \( M(X) \) are compact.

Denote by \( P(X) \) the set of probability measures on \( X \), and by \( P_\leq(X) \) the set of measures \( \mu \) such that \( \mu(X) \leq 1 \). By the Banach–Alaoglu theorem, both \( P(X) \) and \( P_\leq(X) \) are compact and Hausdorff.

**The Riesz pairing** The pairing map

\[ \langle - , - \rangle : C(X) \times M(X) \rightarrow \mathbb{R} \]

is not in general continuous. However, it is continuous on \( C(X) \times P_\leq(X) \):

**Lemma 2.5** Let \( Q \) be a closed bounded subset of \( M(X) \). Then:
i. there is a continuous map $C(X) \to C(Q)$ defined by $f \mapsto \langle f, - \rangle$;

ii. the restriction of the pairing map $C(X) \times M(X) \to \mathbb{R}$ to $C(X) \times Q$ is continuous.

Proof. For (i), first note that for each $f \in C(X)$, the map $\langle f, - \rangle : Q \to \mathbb{R}$ is continuous, by definition of the weak* topology. To show that the resulting map $C(X) \to C(Q)$ is continuous, let $f, g \in C(X)$. Then

$$\|\langle f, - \rangle - \langle g, - \rangle\|_\infty = \sup_{\mu \in Q} |\langle f - g, \mu \rangle| \leq \|f - g\|_\infty \sup_{\mu \in Q} \|\mu\|,$$

and $\sup_{\mu \in Q} \|\mu\|$ is finite as $Q$ is bounded.

Part (ii) follows from Corollary 2.4, since $Q$ is compact (by the Banach–Alaoglu theorem) and Hausdorff.

Supports of functions and measures The support of a function $f : X \to [0, \infty)$ is

$$\text{supp } f = \{x \in X : f(x) > 0\}.$$

Note that we use this set, rather than its closure. Thus, supp $f$ is open when $f$ is continuous.

Every measure $\mu$ on $X$ has a support; that is, there is a smallest closed set $\text{supp } \mu$ such that $\mu(X \setminus \text{supp } \mu) = 0$. (Recall our convention that ‘measure’ means ‘positive Radon measure’, and see, for instance, Chapter III, §2, No. 2 of [3].) The support is characterised by

$$\text{supp } \mu = \{x \in X : \mu(U) > 0 \text{ for all open neighbourhoods } U \text{ of } x\},$$

and has the property that $\int_X f \, d\mu = \int_{\text{supp } \mu} f \, d\mu$ for all $f \in L^1(X, \mu)$.

One of the connections between the two concepts of support is the following.

**Lemma 2.6** Let $\mu$ be a measure on $X$, and let $f : X \to [0, \infty)$ be a continuous function. Then

$$\text{supp } f \cap \text{supp } \mu \neq \emptyset \iff \int_X f \, d\mu > 0.$$

**Proof.** The forwards implication is Proposition 9 in Chapter III, §2, No. 3 of [3], and the backwards implication is trivial.

Approximations to Dirac measures Suppose that we have fixed a measure $\mu$ on our space $X$. The Dirac measure $\delta_x$ at a point $x$ is not in general absolutely continuous with respect to $\mu$, but it can be approximated by absolutely continuous measures, in the following sense:

**Lemma 2.7** Let $\mu$ be a measure on $X$ and $x \in \text{supp } \mu$. For each compact subset $E \subseteq C(\text{supp } \mu)$ and $\varepsilon > 0$, there exists a nonnegative function $u \in C(X)$ such that $u\mu$ is a probability measure and for all $f \in E$,

$$\left| \int_X f \, d(u\mu) - f(x) \right| \leq \varepsilon.$$
Proof. Since the set $E$ of functions on $\text{supp } \mu$ is compact, it is equicontinuous (e.g. by Theorem IV.6.7 of [4]). We can therefore choose a subset $U \subseteq \text{supp } \mu$, containing $x$ and open in $\text{supp } \mu$, such that $|f(y) - f(x)| \leq \varepsilon$ for all $y \in U$ and $f \in E$.

By Urysohn’s lemma we can choose a nonnegative function $u \in C(\text{supp } \mu)$ such that $\text{supp } u \subseteq U$ and $u(x) > 0$. Then $\int_{\text{supp } \mu} u \, d\mu > 0$, so by rescaling we can arrange that $\int_{\text{supp } \mu} u \, d\mu = 1$.

By Tietze’s extension theorem, $u$ can be extended to a nonnegative function continuous on $X$, and then $u\mu$ is a probability measure on $X$. Moreover, for all $f \in E$,

$$\left| \int_X f \, d(u\mu) - f(x) \right| = \left| \int_U (f(y) - f(x))u(y) \, d\mu(y) \right|$$

$$\leq \varepsilon \int_U u(y) \, d\mu(y) = \varepsilon,$$

as required. □

Integral power means Here we review the theory of the power means of a real-valued function on an arbitrary probability space $(X, \mu)$ (now temporarily abandoning the convention that $X$ denotes a compact Hausdorff topological space). This is classical material; for example, Chapter VI of Hardy, Littlewood and Pólya [7] covers the case where $X$ is a real interval.

The essential supremum of a function $f : X \to \mathbb{R}$ with respect to $\mu$ is

$$\text{ess sup}_\mu(f) = \inf \{ a \in \mathbb{R} : \mu(\{ x : f(x) > a \}) = 0 \},$$

and its essential infimum, $\text{ess inf}_\mu(f)$, is defined similarly. A function $f : X \to [0, \infty)$ is essentially bounded if $\text{ess sup}_\mu(f)$ is finite.

Definition 2.8 Let $(X, \mu)$ be a probability space and let $f : X \to [0, \infty)$ be a measurable function such that both $f$ and $1/f$ are essentially bounded. We define for each $t \in [-\infty, \infty]$ a real number

$$M_t(\mu, f) \in (0, \infty),$$

the power mean of $f$ of order $t$, weighted by $\mu$, as follows.

- For $t \in (-\infty, 0) \cup (0, +\infty)$,

$$M_t(\mu, f) = \left( \int_X f^t \, d\mu \right)^{1/t}.$$  \hspace{1cm} (1)

- For $t = 0$,

$$M_0(\mu, f) = \exp \left( \int_X \log f \, d\mu \right).$$
For $t = \pm \infty$,

\[ M_{+\infty}(\mu, f) = \text{ess sup}_\mu f, \]
\[ M_{-\infty}(\mu, f) = \text{ess inf}_\mu f. \]

As we shall see, the definitions in the three exceptional cases are determined by the requirement that the power mean is continuous in its order.

In the case where $X$ is a finite set $\{1, \ldots, n\}$, the definition reduces to that of the finite power means (as in Chapter III of [7]). In particular, the mean of order 0 is the geometric mean:

\[ M_0(\mu, f) = \prod_{i=1}^n f(i)^{\mu(i)}. \]

Remark 2.9 We have assumed here that $f$ and $1/f$ are essentially bounded, or equivalently that

\[ \text{ess inf}_\mu (f) > 0, \quad \text{ess sup}_\mu (f) < \infty. \]

This assumption guarantees that $f^t \in L^1(X, \mu)$ for all real $t$ and that $M_t(\mu, f) \in (0, \infty)$ for all $t \in [-\infty, \infty]$. If $f$ satisfies our assumption then so does $1/f$, and a duality relationship holds:

\[ M_{-t}(\mu, f) = \frac{1}{M_t(\mu, 1/f)}. \]

Power means are increasing and continuous in their order:

**Proposition 2.10** Let $(X, \mu)$ be a probability space and let $f : X \to [0, \infty)$ be a measurable function such that both $f$ and $1/f$ are essentially bounded.

i. If there is some constant $c$ such that $f(x) = c$ for almost all $x \in X$, then $M_t(\mu, f) = c$ for all $t \in [-\infty, \infty]$.

ii. Otherwise, $M_t(\mu, f)$ is strictly increasing in $t \in [-\infty, \infty]$.

**Proof.** Part (i) is trivial. Part (ii) is proved in Section 6.11 of [7] in the case where $X$ is a real interval and $\mu$ is determined by a density function, and the proof extends without substantial change to an arbitrary probability space. \[ \square \]

**Proposition 2.11** Let $(X, \mu)$ be a probability space and let $f : X \to [0, \infty)$ be a measurable function such that both $f$ and $1/f$ are essentially bounded. Then $M_t(\mu, f)$ is continuous in $t \in [-\infty, \infty]$.

**Proof.** Again, this is proved in the case of a real interval in Section 6.11 of [7]. The generalisation to an arbitrary probability space is sketched as Exercise 1.8.1 of [20], although the hypotheses on $f$ there are weaker than ours, and at $t = 0$ only continuity from the right is proved:

\[ \lim_{t \to 0^+} M_t(\mu, f) = M_0(\mu, f). \]

Under our hypotheses on $f$, continuity from the left then follows from the duality stated in Remark 2.9. \[ \square \]
Differentiation under the integral sign  We will require the following standard result in calculus on measure spaces, whose proof can be found in, for example, [9] (Theorem 6.28).

Lemma 2.12 Let \((X, \mu)\) be a measure space and \(J \subseteq \mathbb{R}\) an open interval. Let \(f : X \times J \to \mathbb{R}\) be a map with the following properties:

i. for all \(t \in J\), the map \(f(-, t) : X \to \mathbb{R}\) is integrable;

ii. for almost all \(x \in X\), the map \(f(x, -) : J \to \mathbb{R}\) is differentiable;

iii. there is an integrable function \(h : X \to \mathbb{R}\) such that for all \(t \in J\), for almost all \(x \in X\), we have \(\left|\frac{\partial f}{\partial t}(x, t)\right| \leq h(x)\).

Then \(\frac{\partial f}{\partial t}(-, t) : X \to \mathbb{R}\) is integrable for each \(t \in J\), and the function \(t \mapsto \int_X f(-, t) \mathrm{d}\mu\) is differentiable with derivative \(t \mapsto \int_X \frac{\partial f}{\partial t}(-, t) \mathrm{d}\mu\).

3 Typicality

The setting for the rest of this paper is a space \(X\) equipped with a notion of similarity or proximity between points in \(X\) (which may or may not be derived from a metric). Later, we will study the entropy and diversity of any probability measure on such a space. But first, we show how any probability measure on \(X\) gives rise to a function on \(X\), called its ‘typicality function’, whose value at a point \(x\) indicates how concentrated the measure is near \(x\).

Definition 3.1 Let \(X\) be a compact Hausdorff space. A similarity kernel on \(X\) is a continuous function \(K : X \times X \to [0, \infty)\) satisfying \(K(x, x) > 0\) for all \(x \in X\). The pair \((X, K)\) is a (compact Hausdorff) space with similarities.

Since we will only be interested in compact Hausdorff spaces, we drop the ‘compact Hausdorff’ and simply refer to spaces with similarities. The terminology of similarity originates with the following family of examples.

Example 3.2 There has been vigorous discussion in ecology of how best to quantify the diversity of a biological community. This is a conceptual and mathematical challenge, quite separate from the practical and statistical difficulties, and a plethora of different diversity measures have been proposed over 70 years of debate in the ecological literature [16].

Any realistic diversity measure should reflect the degree of variation between the species present. All else being equal, a lake containing four species of carp should be counted as less diverse than a lake containing four very different species of fish. The similarity between species may be measured genetically, phylogenetically, functionally, or in some other way (as discussed in [13]); how it is done will not concern us here.

Mathematically, we take a finite set \(X = \{1, \ldots, n\}\) (whose elements represent the species) and a real number \(Z_{ij} \geq 0\) for each pair \((i, j)\) (representing the degree of similarity between species \(i\) and \(j\)). A similarity coefficient \(Z_{ij} = 0\)
means that species $i$ and $j$ are completely dissimilar, and we therefore assume that $Z_{ii} > 0$ for all $i$. Thus, $Z = (Z_{ij})$ is an $n \times n$ nonnegative real matrix with strictly positive entries on the diagonal.

Many ways of assigning inter-species similarities are calibrated on a scale of 0 to 1, with $Z_{ii} = 1$ for all $i$ (each species is identical to itself). For example, percentage genetic similarity gives similarity coefficients in $[0, 1]$, as does the similarity measure $e^{-d(i,j)}$ derived from a metric $d$ and discussed below. The simplest possible choice of $Z$ is the identity matrix, embodying the crude assumption that different species have nothing whatsoever in common.

In the language of Definition 3.1, we are considering here the case of finite spaces with similarities: $X = \{1, \ldots, n\}$ (with the discrete topology) and the similarity kernel $K$ is given by $K(i, j) = Z_{ij}$. When $Z$ is the identity matrix, $K$ is the Kronecker delta.

**Example 3.3** Any compact metric space $(X, d)$ can be regarded as a space with similarities $(X, K)$ by putting

$$K(x, y) = e^{-d(x, y)}$$

$(x, y \in X)$. The extreme case where $d(x, y) = \infty$ for all $x \neq y$ produces the Kronecker delta.

Although the negative exponential is not the only reasonable function transforming distances into similarities, it turns out to be a particularly fruitful choice. It is associated with the very fertile theory of the magnitude of metric spaces (surveyed in [15]), which has deep connections with convex geometry, geometric measure theory and potential theory. Moreover, the general categorical framework of magnitude all but forces this choice of transformation, as explained in Example 2.4(3) of [15].

In the examples above, the similarity kernel is symmetric: $K(x, y) = K(y, x)$ for all $x, y \in X$. We do not include symmetry in the definition of similarity kernel, partly because asymmetric similarity matrices occasionally arise in ecology (e.g. [13], Appendix, Proposition A7), and also because of the argument of Gromov ([6], p. xv) and Lawvere ([10], p. 138–9) that the symmetry condition in the definition of metric can be too restrictive. To obtain our main result, however, it will be necessary to add symmetry as a hypothesis on $K$.

Most measures of biological diversity depend (at least in part) on the relative abundance distribution $p = (p_1, \ldots, p_n)$ of the species, where ‘relative’ means that the $p_i$ are normalised to sum to 1. Multiplying the similarity matrix $Z$ by the column vector $p$ gives another vector $Zp$, with $i$th entry

$$(Zp)_i = \sum_j Z_{ij}p_j.$$  

This is the expected similarity between an individual of species $i$ and an individual chosen at random. Thus, $(Zp)_i$ measures how typical individuals of species $i$ are within the community. The generalisation to an arbitrary space with similarities is as follows.
**Definition 3.4** Let \((X, K)\) be a space with similarities. For each \(\mu \in M(X)\) and \(x \in X\), define
\[
(K\mu)(x) = \int_X K(x, -) \, d\mu \in \mathbb{R}.
\]
This defines a function \(K\mu : X \rightarrow \mathbb{R}\), the **typicality function** of \((X, K, \mu)\).

When \(\mu\) is a probability measure (the case of principal interest), \((K\mu)(x)\) is the expected similarity between \(x\) and a random point. It therefore detects the extent to which \(x\) is similar, or near, to sets of large measure.

In the next section, we will define entropy and diversity in terms of the typicality function \(K\mu\). For that, we will need \(K\mu\) to satisfy some analytic conditions, which we establish now.

**For the rest of this section, let \((X, K)\) be a space with similarities.**

**Lemma 3.5** The function \(\overline{K} : X \rightarrow C(X)\) defined by \(x \mapsto K(x, -)\) is continuous.

**Proof.** Since \(X\) is compact Hausdorff and \(K\) is continuous, this follows from Corollary 2.4.

**Lemma 3.6** For each \(\mu \in M(X)\), the function \(K\mu : X \rightarrow \mathbb{R}\) is continuous.

**Proof.** Note that \(K\mu\) is the composite
\[
X \xrightarrow{\overline{K}} C(X) \xrightarrow{(-,\mu)} \mathbb{R}.
\]
We have just proved that \(\overline{K}\) is continuous, and \((-,\mu) = \int_X - \, d\mu\) is a continuous linear functional. Hence \(K\mu\) is continuous.

**Lemma 3.7** The map
\[
K_* : P(X) \rightarrow C(X) \quad \mu \mapsto K\mu
\]
is continuous.

**Proof.** Both \(X\) and \(P(X)\) are compact Hausdorff so, applying Corollary 2.4 twice, an equivalent statement is that the map
\[
X \rightarrow C(P(X)) \quad x \mapsto (K\mu)(x) = \langle K\mu, x \rangle = \int_X K(x, -) \, d\mu
\]
is continuous. This map is the composite
\[
X \xrightarrow{\overline{K}} C(X) \xrightarrow{\langle -, \mu \rangle} C(P(X)),
\]
where the second map is \(f \mapsto \langle f, - \rangle\) and is continuous by Lemma 2.5(i). Hence, \(K_* : P(X) \rightarrow C(X)\) is continuous.
Proposition 3.8 For every measure \( \mu \) on \( X \), the typicality function \( K \mu \) has the following properties:

i. \( \text{supp} \, K \mu \supseteq \text{supp} \, \mu \).

ii. Both \( K \mu \) and \( 1/K \mu \) are essentially bounded with respect to \( \mu \).

Proof. For (i), let \( x \in \text{supp} \, \mu \). Since \( K \) is positive on the diagonal, \( x \in \text{supp} \, K(x,-) \), so \( \text{supp} \, \mu \cap \text{supp} \, K(x,-) \neq \emptyset \). Hence by Lemma 2.6,

\[
(K\mu)(x) = \int_X K(x,-) \, d\mu > 0.
\]

For (ii), \( \text{supp} \, \mu \) is compact, and \( K \mu \) is continuous with \( K \mu \big|_{\text{supp} \, \mu} > 0 \), so both \( K \mu \) and \( 1/K \mu \) are bounded on \( \text{supp} \, \mu \). Hence both are essentially bounded on \( X \). \( \Box \)

4 Diversity and entropy

Here we introduce the main subject of the paper: a one-parameter family of functions that quantify the degree of spread of a probability measure on a compact Hausdorff space \( X \), with respect to a chosen notion of similarity between points of \( X \).

Specifically, take a probability measure \( \mu \) on a space with similarities \((X,K)\). The measure can be regarded as widely spread across \( X \) if most points are dissimilar to most of the rest of \( X \), or in other words, if the typicality function \( K \mu : X \to \mathbb{R} \) takes small values on most of \( X \). An equivalent way to say this is that the ‘atypicality’ function \( 1/K \mu \) takes large values on most of \( X \). In ecological terms, a community is diverse if it is predominantly made up of species that are unusual or atypical within that community (for example, many rare and highly dissimilar species).

The diversity of \( \mu \) is, therefore, defined as the mean atypicality of a point. It is useful to consider not just the arithmetic mean, but the power means of all orders:

**Definition 4.1** Let \((X,K)\) be a space with similarities and let \( q \in [-\infty, \infty] \).

The **diversity of order** \( q \) of a probability measure \( \mu \) on \( X \) is

\[
D^K_q(\mu) = M_{1-q}(\mu, 1/K \mu) \in (0, \infty).
\]

The **entropy of order** \( q \) of \( \mu \) is \( H^K_q(\mu) = \log D^K_q(\mu) \).

By the duality of Remark 2.9, an equivalent definition is

\[
D^K_q(\mu) = 1/M_{q-1}(\mu, K \mu).
\]

On the right-hand side, the denominator is the mean typicality of a point in \( X \), which is a measure of lack of diversity; its reciprocal is then a measure of diversity. The power means in this formula and Definition 4.1 are well-defined because \( K \mu \) and \( 1/K \mu \) are essentially bounded with respect to \( \mu \) (Proposition 3.8).
Explicitly,

\[
D^K_q(\mu) = \begin{cases} 
\left( \int_X (K\mu)^{q-1} \, d\mu \right)^{1/(1-q)} & \text{if } q \in (-\infty, 1) \cup (1, \infty), \\
\exp \left( -\int_X \log(K\mu) \, d\mu \right) & \text{if } q = 1, \\
\frac{1}{\text{ess sup}_\mu K\mu} & \text{if } q = \infty, \\
\frac{1}{\text{ess inf}_\mu K\mu} & \text{if } q = -\infty.
\end{cases}
\]

We usually work with the diversities $D^K_q$ rather than the entropies $H^K_q$, but evidently it is trivial to translate results on diversity into results on entropy.

**Example 4.2** Let $X$ be the set $\{1, \ldots, n\}$ with the discrete topology, let $K$ be the Kronecker delta on $X$ (the ‘simplest possible choice’ of Example 3.2), and let $\mu$ be the uniform measure on $X$. Then $K\mu \equiv 1/n$, so $D^K_q(\mu) = n$ and $H^K_q(\mu) = \log n$ for all $q$. This conforms to the intuition that the larger we take $n$ to be, the more thinly spread the uniform measure on $\{1, \ldots, n\}$ becomes.

The next two examples also concern the finite case. They are described in terms of the ecological scenario of Example 3.2. Thus, $X = \{1, \ldots, n\}$ is a set of species, $Z_{ij} = K(i, j)$ is the similarity between species $i$ and $j$, and $\mu = p = (p_1, \ldots, p_n)$ gives the proportions in which the species are present.

**Example 4.3** Put $Z = I$ (distinct species have nothing in common). Then the diversity of order 0 is

\[
D_I^0(p) = \sum_{i \in \text{supp } p} p_i \cdot \frac{1}{p_i} = |\text{supp } p|.
\]

This is just the number of species present. It is the simplest diversity measure of all. But it takes no account of species abundances beyond presence and absence, whereas ordinarily, for instance, a community of two species is considered more diverse if they are equally abundant than if their proportions are $(0.99, 0.01)$.

The diversities of nonzero orders do, however, reflect the balance between species. For example, the diversity of order 1 is

\[
D_I^1(p) = \exp \left( -\sum_{i \in \text{supp } p} p_i \log p_i \right) = \prod_{i \in \text{supp } p} p_i^{-p_i}
\]

and the entropy $H_I^1(p) = \log D_I^1(p)$ of order 1 is the Shannon entropy $-\sum p_i \log p_i$, which can be understood as measuring the uniformity of the distribution $p$. The diversity of order 2 is

\[
D_I^2(p) = \frac{1}{\sum_{i=1}^n p_i^2}.
\]

The denominator is the probability that two individuals are chosen at random are of the same species, and $D_I^2(p)$ itself is the expected number of such trials needed in order to obtain a matching pair. The diversity of order $\infty$ is

\[
D_I^\infty(p) = \frac{1}{\max_i p_i},
\]
which measures the extent to which the community is dominated by a single species. All four of these diversity measures (or simple transformations of them) are used by ecologists [16]. For a general parameter value \( q \neq 1, \pm \infty \), the diversity of order \( q \) is

\[
D^I_q(p) = \left( \sum_{i \in \text{supp} p} p_i^q \right)^{1/(1-q)}.
\]

In ecology, \( D^I_q \) is known as the Hill number of order \( q \) [8], and in information theory, \( H^I_q = \log D^I_q \) is called the Rényi entropy of order \( q \) [22]. For reasons explained in Remark 6.1, we usually restrict to \( q \geq 0 \).

The parameter \( q \) controls the emphasis placed on rare or common species. Low values of \( q \) emphasise rare species; high values emphasise common species. At one extreme, \( D^I_0(p) \) is simply the number of species present, regardless of abundance; thus, diversity of order 0 attaches as much importance to rare species as common ones. At the other, diversity of order \( \infty \) depends only on the abundance of the most common species, completely ignoring rarer ones.

If a community loses one or more rare species, while at the same time the species that remain become more evenly balanced, its low-order diversity will fall but its high-order diversity will rise. For example, \( D^I_q \) measures the relative abundance distribution \((0.5,0.5,0)\) as less diverse than \((0.8,0.1,0.1)\) when \( q < 0.852 \ldots \), but more diverse for all higher values of \( q \).

The moral is that when judging which of two communities is the more diverse, the answer depends critically on the parameter \( q \). Different values of \( q \) may produce opposite judgements.

**Example 4.4** Still in the ecological setting, consider now a general similarity matrix \( Z \), thus taking into account the varying similarities between species (as in Example 3.2). The diversity measures \( D^Z_q \) and the role of the parameter \( q \) can be understood much as in the case \( Z = I \), but now incorporating inter-species similarity. For example,

\[
D^Z_2(p) = 1/\sum_{i,j} p_i Z_{ij} p_j
\]

is the reciprocal expected similarity between a random pair of individuals (rather than the reciprocal probability that they are of the same species), and

\[
D^Z_\infty(p) = 1/\max_{i \in \text{supp} p} (Zp)_i
\]

reflects the dominance of the largest cluster of species (rather than merely the largest single species).

**Example 4.5** Let \((X,K)\) be an arbitrary space with similarities. Among all the diversity measures \((D^K_q)_{q \in [0,\infty]} \), one with especially convenient mathematical properties is the diversity of order 2:

\[
D^K_2(\mu) = \frac{1}{\int_X \int_X K(x,y) d\mu(x) d\mu(y)}.
\]
Meckes used $D^K_2$, and more particularly the maximum diversity 
\[ \sup_{\mu \in P(X)} D^K_2(\mu) \] of order 2, to prove results on the Minkowski dimension of metric spaces ([17], Section 7). These include not only Theorem 1.1, but also results that do not mention maximum diversity in their statements.

We now establish the basic analytic properties of diversity. First, we show that for a fixed probability measure $\mu$, the diversity $D^K_q(\mu)$ is a continuous and decreasing function of its order $q$. Second, we show that for fixed $q \in (0, \infty)$, the diversity $D^K_q(\mu)$ is continuous in the measure $\mu$. The first fact is immediate, but proving the second is considerably more involved.

**Proposition 4.6** Let $(X, K)$ be a space with similarities and let $\mu \in P(X)$.

i. $D^K_q(\mu)$ is continuous in its order $q \in [-\infty, \infty]$.

ii. If $K\mu$ is constant on the support of $\mu$, then the function $q \mapsto D^K_q(\mu)$ is constant on $[-\infty, \infty]$; otherwise, it is strictly decreasing in $q \in [-\infty, \infty]$.

**Proof.** Part (i) follows from Proposition 2.11, and part (ii) from Proposition 2.10. \qed

**Remark 4.7** A critical role will be played by measures $\mu$ satisfying the first case of Proposition 4.6(ii). We call $\mu$ **balanced** if the function $K\mu$ is constant on $\text{supp} \mu$. (In [14], where $X$ is taken to be finite, such measures were called ‘invariant’.) Equivalently, $\mu$ is balanced if $D^K_q(\mu)$ is constant over $q \in [-\infty, \infty]$.

If $(K\mu)|_{\text{supp} \mu}$ has constant value $c$ then $D^K_q(\mu)$ has constant value $1/c$.

**Proposition 4.8** Let $(X, K)$ be a space with similarities. For every $q \in (0, \infty)$, the diversity function $D^K_q : P(X) \to \mathbb{R}$ is continuous.

(Recall that we always use the weak* topology on $P(X)$.)

The proof of this proposition takes the form of three lemmas, addressing the three cases $q \in (1, \infty)$, $q \in (0, 1)$ and $q = 1$. In the first case, matters are straightforward.

**Lemma 4.9** For every $q \in (1, \infty)$, the diversity function $D^K_q : P(X) \to \mathbb{R}$ is continuous.

**Proof.** The map $D^K_q$ is the composite

\[
P(X) \xrightarrow{\Delta} P(X) \times P(X) \xrightarrow{K^*_\times \text{Id}} C(X) \times P(X) \xrightarrow{(-)^{q-1} \times \text{Id}} C(X) \times P(X) \xrightarrow{(-, -)} \mathbb{R} \xrightarrow{(-)^{1/(q-1)}} \mathbb{R}.
\]

Here $\Delta$ is the diagonal, which is certainly continuous. The map $K^*_\times \text{Id}$ was defined and proved to be continuous in Lemma 3.7, and $(-)^{q-1} : C(X) \to C(X)$ is continuous by Lemma 2.3. The restricted pairing $(-, -)$ on $C(X) \times P(X)$ is continuous by Lemma 2.5. Finally, $(-)^{1/(q-1)}$ is evidently continuous. Hence $D^K_q$ is continuous. \qed
The case $q \in (0, 1)$ is more difficult. The main work goes into handling the possibility that $(K\mu)(x) = 0$ for some $x$, in which case the function $(K\mu)^{q-1}$ is not defined everywhere.

Our strategy is as follows. Were we to assume that $K(x, y) > 0$ for all $x, y \in X$, then $K\mu$ would be positive everywhere, by compactness. We would then be free of the subtlety just described.

We approximate this convenient situation by covering $X$ with subsets $U_1, \ldots, U_n$ such that $K$ is strictly positive on $U_i \times U_i$, for each $i$. (That is, points within each $U_i$ are reasonably similar.) We then decompose the function $(D^K_q)^{1-q}$ as a sum of functions $d_i$. Roughly speaking (and we will make an accurate statement shortly), $d_i(\mu)$ is the contribution

$$\int_{U_i} (K\mu)^{q-1} \, d\mu$$

to $(D^K_q)^{1-q}$ made on $U_i$. It is enough to show that each of these functions $d_i : P(X) \to \mathbb{R}$ is continuous.

The most delicate part of the argument is to show that $d_i$ is continuous at measures $\mu \in P(X)$ whose support does not meet $U_i$. This is because although the integral (2) vanishes at $\mu$, there may be measures $\nu \in P(X)$ arbitrarily close to $\mu$ whose support does meet $U_i$. In that case, the integral (2) for $\nu$ does not vanish, and the function $K\nu$ may take small values on $U_i$, in which case the integrand $(K\nu)^{q-1}$ is large. But in order for $d_i$ to be continuous, we need the integral of this large function to be small. The details of the argument involve an estimate that depends on the particular form of the diversity function.

The proof below implements the argument just sketched, with one further refinement. In keeping with the viewpoint of $M(X)$ as the dual of $C(X)$, we primarily work not with the cover $(U_i)$ itself, but with a continuous partition of unity $(p_i)$ subordinate to it. The formula (2) for $d_i(\mu)$ is adapted accordingly, effectively replacing the characteristic function of $U_i$ by $p_i$.

**Lemma 4.10** For every $q \in (0, 1)$, the diversity function $D^K_q : P(X) \to \mathbb{R}$ is continuous.

**Proof.** This proof proceeds in four steps.

**Step 1: partitioning the space** Put

$$b = \frac{1}{2} \inf_{x \in X} K(x, x).$$

By the topological hypotheses, $b > 0$ and we can find a finite open cover $U_1, \ldots, U_n$ of $X$ such that $K(x, y) \geq b$ whenever $x, y \in U_i$ for some $i$. We can also find a continuous partition of unity $p_1, \ldots, p_n$ such that $\text{supp} \, p_i \subseteq U_i$ for each $i$.

For all $\mu \in P(X)$,

$$D^K_q(\mu)^{1-q} = \int_X (K\mu)^{q-1} \, d\mu = \sum_{i=1}^{n} \int_X (K\mu)^{q-1} p_i \, d\mu.$$
To see that $D^K_q$ is continuous it will suffice to show that, for each $i$, the map

$$d_i : P(X) \to \mathbb{R}$$

$$\mu \mapsto \int_X (K\mu)^q p_i \, d\mu$$

is continuous. For the rest of the proof, fix $i \in \{1, \ldots, n\}$.

**Step 2: bounding $K\mu$ below** Let $\mu \in P(X)$. Then for all $x \in U_i$,

$$(K\mu)(x) = \int_X K(x, y) \, d\mu(y)$$

$$\geq \int_{U_i} K(x, y)p_i(y) \, d\mu(y)$$

$$\geq b \int_X p_i \, d\mu.$$

Thus, $(K\mu)(x) \geq b \int p_i \, d\mu$ for all $x \in U_i$. By Lemma 2.6, this lower bound on $(K\mu)$ is strictly positive if $\text{supp } p_i \cap \text{supp } \mu \neq \emptyset$.

**Step 3: continuity at nontrivial measures** Here we show that $d_i$ is continuous at each element of the set

$$P_i(X) = \{\mu \in P(X) : \text{supp } p_i \cap \text{supp } \mu \neq \emptyset\}.$$

Lemma 2.6 implies that $\mu \in P_i(X)$ if and only if $\int p_i \, d\mu > 0$, so $P_i(X)$ is an open subset of $P(X)$. Thus, it is equivalent to show that the restriction of $d_i$ to $P_i(X)$ is continuous.

We begin by showing that there is a well-defined, continuous map $G_i : P_i(X) \to C(U_i)$ given by

$$G_i(\mu) = (K\mu)^{q-1} |_{U_i}.$$

It is well-defined because, for each $\mu$, the map $K\mu$ is continuous and strictly positive on $U_i$ (by step 2). To show that $G_i$ is continuous, consider the following spaces and maps, defined below:

$$P_i(X) \xrightarrow{K_*} C_i^+(X) \xrightarrow{\text{res}} C^+(U_i) \xrightarrow{(-)^{q-1}} C^+(U_i) \hookrightarrow C(U_i).$$

Here

$$C_i^+(X) = \{f \in C(X) : f(x) > 0 \text{ for all } x \in U_i\},$$

$$C^+(U_i) = \{g \in C(U_i) : g(x) > 0 \text{ for all } x \in U_i\} = \text{Top}(U_i, (0, \infty)).$$

The first map $K_*$ is the restriction of $K_* : P(X) \to C(X)$; the restricted $K_*$ is well-defined by step 2 and continuous by Lemma 3.7. The second map is
restriction, which is certainly continuous, the third map \((-)^{q-1}\) is continuous by Lemma 2.3 and compactness of \(U_i\), and the last map is inclusion. The composite of these maps is \(G_i\), which is therefore also continuous, as claimed.

To show that \(d_i\) is continuous on \(P_i(X)\), consider the chain of maps

\[
P_i(X) \xrightarrow{\triangle} P_i(X) \times P(X) \xrightarrow{G_i \times (p_i \cdot \cdot)} C(U_i) \times P(U_i) \xrightarrow{\cdot \cdot} \mathbb{R}.
\]

The first map is the diagonal followed by an inclusion; it is continuous. In the second, \(p_i \cdot \cdot\) is a restriction of the linear map \(M(X) \to M(U_i)\) defined by \(\mu \mapsto p_i \mu\), which is continuous. Since \(G_i\) is continuous too, so is \(G_i \times (p_i \cdot \cdot)\). The third map is continuous by Lemma 2.5(ii). And the composite of the chain is \(d_i|_{P_i(X)}\), which is, therefore, also continuous.

**Step 4: continuity at trivial measures** Finally, we show that the function \(d_i\) is continuous at all points \(\mu \in P(X)\) such that \(\text{supp} p_i \cap \text{supp} \mu = \emptyset\). Fix such a \(\mu\).

Given \(\nu \in P(X)\), if \(\text{supp} p_i \cap \text{supp} \nu = \emptyset\) then \(d_i(\nu) = 0\), and otherwise

\[
d_i(\nu) = \int_{U_i} (K\nu)^{q-1} p_i \, d\nu \leq \int_{U_i} \left( b \int_X p_i \, d\nu \right)^{q-1} p_i \, d\nu = b^{q-1} \left( \int_X p_i \, d\nu \right)^q
\]

(using step 2 and that \(q < 1\)). So in either case,

\[
0 \leq d_i(\nu) \leq b^{q-1} \left( \int_X p_i \, d\nu \right)^q. \tag{3}
\]

Now as \(\nu \to \mu\) in \(P(X)\), we have

\[
\int_X p_i \, d\nu \to \int_X p_i \, d\mu = 0,
\]

so

\[
b^{q-1} \left( \int_X p_i \, d\nu \right)^q \to 0
\]

(using that \(q > 0\)). Hence the bounds (3) give \(d_i(\nu) \to 0 = d_i(\mu)\), as required.

The remaining case, \(q = 1\), will be deduced from the cases \(q \in (0, 1)\) and \(q \in (1, \infty)\), exploiting the fact that \(D^K_q(\mu)\) is decreasing in \(q\).

**Lemma 4.11** The diversity function \(D^K_1 : P(X) \to \mathbb{R}\) is continuous.

**Proof.** Let \(\mu \in P(X)\) and \(\varepsilon > 0\). Since \(D^K_q(\mu)\) is continuous and decreasing in \(q\) (Proposition 4.6), we can choose \(q^+ \in (1, \infty)\) such that

\[
0 \leq D^K_1(\mu) - D^K_{q^+}(\mu) < \varepsilon/2.
\]
Since $D^K_q : P(X) \to \mathbb{R}$ is continuous, we can find a neighbourhood $U^+$ of $\mu$ such that for all $\nu \in U^+$,

$$|D^K_{q^+}(\mu) - D^K_{q^+}(\nu)| < \varepsilon/2.$$ 

Then for all $\nu \in U^+$,

$$D^K_1(\nu) \geq D^K_{q^+}(\nu) \geq D^K_1(\mu) - \varepsilon.$$

Similarly, we can find a neighbourhood $U^-$ of $\mu$ such that for all $\nu \in U^-$,

$$D^K_1(\nu) \leq D^K_1(\mu) + \varepsilon$$

Hence $|D^K_1(\nu) - D^K_1(\mu)| \leq \varepsilon$ for all $\nu \in U^+ \cap U^-$. \hfill $\Box$

This completes the proof of Proposition 4.8: the diversity function of each finite positive order is continuous.

**Remark 4.12** Proposition 4.8 excludes the cases $q = 0$ and $q = \infty$. Diversity of order 0 is not continuous even in the simplest case of a finite set and the identity similarity matrix; for as we saw in Example 4.3, $D^0_I(p)$ is the cardinality of $\text{supp } p$, which is not continuous in $p$. Diversity of order $\infty$ need not be continuous either. For example, take $X = \{1, 2, 3\}$ and the similarity matrix

$$Z = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

and put $p = (1/2 - t, 2t, 1/2 - t)$. Then $D^Z_{\infty}(p)$ is 1 if $t \in (0, 1/2)$, but 2 if $t = 0$.

## 5 Magnitude

In order to show that maximum diversity and maximum entropy are well-defined, we first have to define a closely related invariant, magnitude. Magnitude has been defined and studied at various levels of generality, including finite enriched categories and compact metric spaces, for which it has strong geometric content. (See [15] for a survey.) We will define the magnitude of a space with similarities.

First we consider signed measures whose typicality function takes constant value 1.

**Definition 5.1** Let $X = (X, K)$ be a space with similarities. A **weight measure** on $X$ is a signed measure $\mu \in M(X)$ such that $K\mu \equiv 1$ on $X$.

This generalises the definition of weight measure on a compact metric space, introduced in [28]. Note that despite our convention that ‘measure’ means positive measure, a weight measure is a **signed** measure.
Example 5.2 Let \( X = \{1, \ldots, n\} \) be a finite set, writing, as usual, \( K(i, j) = Z_{ij} \). Then a weight measure on \( X \) is a vector \( w \in \mathbb{R}^n \) such that \( (Zw)_i = 1 \) for \( i = 1, \ldots, n \). If \( Z \) is invertible then there is exactly one weight measure, but in general there may be none or many. Even if \( Z \) has many weight measures, the total weight \( \sum_i w_i \) turns out to be independent of the weighting \( w \) chosen, just as long as \( Z \) is symmetric (or, more generally, the transpose of \( Z \) admits a weighting too). This common quantity \( \sum_i w_i \) is called the magnitude of \( (X, K) \), and its independence of the weighting chosen is a special case of the following result.

A space with similarities \( (X, K) \) is symmetric if \( K \) is symmetric.

Lemma 5.3 Let \( (X, K) \) be a symmetric space with similarities. Then \( \mu(X) = \nu(X) \) for any weight measures \( \mu \) and \( \nu \) on \( X \).

Proof. Since \( \nu \) is a weight measure,
\[
\mu(X) = \int_X d\mu(x) = \int_X \left( \int_X K(x, y) d\nu(y) \right) d\mu(x).
\]
Since \( \mu \) is a weight measure,
\[
\nu(X) = \int_X d\nu(y) = \int_X \left( \int_X K(y, x) d\mu(x) \right) d\nu(y).
\]
So by symmetry of \( K \) and Tonelli’s theorem, \( \mu(X) = \nu(X) \).

This lemma makes the following definition valid.

Definition 5.4 Let \( (X, K) \) be a symmetric space with similarities admitting at least one weight measure. The magnitude of \( (X, K) \) is
\[
|\!(X, K)\!| = \mu(X),
\]
for any weight measure \( \mu \) on \( (X, K) \). We often write \(|\!(X, K)\!|\) as just \(|X|\).

We will mostly be concerned with positive weight measures. (Note that in an unfortunate clash of terminology, a weight measure on a finite set is positive if and only if the corresponding vector is nonnegative.)

Lemma 5.5 Let \( (X, K) \) be a symmetric space with similarities admitting a positive weight measure. Then \(|X| \geq 0\), with equality if and only if \( X = \emptyset \).

Proof. The inequality is immediate from the definition of magnitude, as is the fact that the magnitude of the empty set is 0. Now suppose that \( X \) is nonempty. Choose \( x \in X \) and a positive weight measure \( \mu \) on \( (X, K) \). Since \( \int_X K(x, -) d\mu = 1 \), the measure \( \mu \) is nonzero. Hence, \(|X| = \mu(X) > 0\).

Let \( (X, K) \) be a space with similarities. Given a closed subset \( Y \) of \( X \), we regard \( Y \) as a space with similarities by restriction of the similarity kernel \( K \).
Any measure \( \nu \neq 0 \) on \( Y \) can be normalised and extended by zero to give a probability measure \( \hat{\nu} \) on \( X \), defined by

\[
\hat{\nu}(U) = \frac{\nu(U \cap Y)}{\nu(Y)}
\]

for all Borel sets \( U \subseteq X \). In particular, whenever \( Y \neq \emptyset \) and \( \nu \) is a positive weight measure on \( Y \), the probability measure \( \hat{\nu} \) on \( X \) is well-defined (by Lemma 5.5), with

\[
\hat{\nu}(U) = \frac{\nu(U \cap Y)}{|Y|}
\]

for all Borel sets \( U \subseteq X \). The construction \( \nu \mapsto \hat{\nu} \) relates the notion of weight measure to that of balanced measure (Remark 4.7) as follows.

**Lemma 5.6** Let \((X, K)\) be a symmetric space with similarities. The following are equivalent for a probability measure \( \mu \) on \( X \):

i. \( \mu \) is balanced (that is, \( K \mu \) is constant on \( \text{supp} \mu \));

ii. the function \( q \mapsto D^K_q(\mu) \) is constant on \([-\infty, \infty]\);

iii. \( \mu = \hat{\nu} \) for some positive weight measure \( \nu \) on \( \text{supp} \mu \);

iv. \( \mu = \hat{\nu} \) for some positive weight measure \( \nu \) on some nonempty closed subset \( Y \subseteq X \).

When these conditions hold, \( D^K_q(\mu) = |Y| \) for all nonempty \( Y \subseteq X \) admitting a positive weight measure \( \nu \) such that \( \hat{\nu} = \mu \), and all \( q \in [-\infty, \infty] \).

**Proof.** The equivalence of (i) and (ii) is Proposition 4.6(ii).

Now assuming (i), we prove (iii). Write \( c \) for the constant value of \( K \mu \) on \( \text{supp} \mu \). Then \( c > 0 \) by Proposition 3.8(i), so we can define a measure \( \nu \) on \( \text{supp} \mu \) by \( \nu(W) = \mu(W)/c \) for all Borel sets \( W \subseteq \text{supp} \mu \). Then \( \nu \) is a weight measure on \( \text{supp} \mu \), since for all \( y \in \text{supp} \mu \),

\[
(K\nu)(y) = \int_{\text{supp} \mu} K(y, -) \, d\nu = \frac{1}{c} \int_X K(y, -) \, d\mu = \frac{1}{c} (K\mu)(y) = 1.
\]

Finally, \( \hat{\nu} = \mu \): for given a Borel set \( U \subseteq X \),

\[
\hat{\nu}(U) = \frac{\nu(U \cap \text{supp} \mu)}{\nu(\text{supp} \mu)} = \frac{\mu(U \cap \text{supp} \mu)}{\mu(\text{supp} \mu)} = \mu(U),
\]

since \( \mu \) is a probability measure. This completes the proof that (i) implies (iii).

Trivially, (iii) implies (iv). Finally, we assume (iv) and prove (i). Take \( Y \) and \( \nu \) as in (iv). For all \( x \in \text{supp} \mu \),

\[
(K\nu)(x) = \int_X K(x, -) \, d\nu = \frac{1}{\nu(Y)} \int_Y K(x, -) \, d\nu = \frac{1}{\nu(Y)},
\]

since \( \nu \) is a weight measure on \( Y \). This proves (i). It also proves the final statement: for by Remark 4.7,

\[
D^K_q(\mu) = \nu(Y) = |Y|
\]

for all \( q \in [-\infty, \infty] \). \( \square \)
6 Balanced and maximising measures

In the case of the Kronecker delta on a finite discrete space, maximising diversity is very simple. Indeed, it is a classical and elementary result that for each \( q \in [0, \infty] \), the Rényi entropy \( H_q^I \) of order \( q \) (Example 4.3) is maximised by the uniform distribution, and that unless \( q = 0 \), the uniform distribution is unique with this property. The same is therefore true of the diversity measures \( D_q^I \).

For a finite space with an arbitrary similarity kernel, maximising measures are no longer uniform \([14]\). We cannot, therefore, expect that on a general space with similarities, diversity is maximised by the ‘uniform’ measure (whatever that might mean). Nevertheless, maximising measures have a different uniformity property: they are balanced. That is the main result of this section.

Remark 6.1 We usually restrict the parameter \( q \) to lie in the range \([0, \infty]\). Even in the simplest case of the Kronecker delta on a finite set, \( D_q^K \) and \( H_q^K \) behave quite differently for negative \( q \) than for positive \( q \). When \( q < 0 \), the uniform measure no longer maximises \( D_q^K \) or \( H_q^K \), and in fact minimises them among all measures of full support (as can be shown using Proposition 4.6(ii)).

For the rest of this section, let \((X, K)\) be a symmetric space with similarities.

Definition 6.2 Let \( q \in [0, \infty] \). A probability measure on \( X \) is \( q \)-maximising if it maximises \( D_q^K \). A probability measure on \( X \) is maximising if it is \( q \)-maximising for all \( q \in [0, \infty] \).

We will show in Section 7 that any measure that is \( q \)-maximising for some \( q > 0 \) is, in fact, maximising. That proof will depend on the next result: that any measure that is \( q \)-maximising for some \( q \in (0, 1) \) is balanced.

We prove this using a variational argument. The strategy is similar to that used in the finite case ([14], Section 5), which can be interpreted as follows.

Let \( X \) be a set of species with symmetric similarity matrix \( Z \), and let \( p \) be a probability distribution on \( X \). Take \( i^-, i^+ \in \text{supp} p \) with the property that \((Zp)_{i^-}\) is minimal and \((Zp)_{i^+}\) is maximal. Ecologically, then, \( i^+ \) is the least typical species present in the community, and \( i^- \) is the most typical.

Suppose that \( p \) is not balanced, that is, some species are more typical than others. Then \((Zp)_{i^-} < (Zp)_{i^+}\). For \( t \in \mathbb{R} \), write

\[
p_t = p + t(\delta_{i^-} - \delta_{i^+}),
\]

where \( \delta_i \) is the vector with \( i^{\text{th}} \) entry equal to 1 and all other entries 0. When \( t \) is sufficiently small, \( p_t \) is a probability distribution on \( X \), and describes the relative abundance distribution after \( t \) units of the most typical species \( i^+ \) have been replaced by \( t \) units of the least typical species \( i^- \). It is intuitively plausible that this substitution should increase the diversity of the community, and indeed it can be shown that, at least for \( q \in (0, 1) \), the derivative of \( D_q^K(p_t) \) at \( t = 0 \) is strictly positive.
Now fix $q \in (0, 1)$, and suppose that $p$ is $q$-maximising. Then the derivative of $D^K_q(p_t)$ at $t = 0$ is 0. Hence, by the previous paragraph, the distribution $p$ is balanced. The moral is that although a $q$-maximising distribution does not generally have all species equally abundant, they are all equally typical.

Extending the argument from finite spaces to compact spaces introduces complications of two kinds. First, there are routine matters arising from replacing sums by integrals. But more significantly, and unlike in the finite case, the Dirac measure $\delta_x$ need not be absolutely continuous with respect to $\mu$ for points $x$ in the support of a probability measure $\mu$. For this reason, when $x^\pm$ are the least and most typical points of $X$, the signed measure $\mu + t(\delta_{x^-} - \delta_{x^+})$ need not be positive (hence, is not a probability measure), even for small $t$. We therefore use approximations to Dirac measures, as provided by Lemma 2.7.

**Proposition 6.3** For $q \in (0, 1)$, every $q$-maximising measure on $(X, K)$ is balanced.

**Proof.** Let $q \in (0, 1)$ and let $\mu$ be a $q$-maximising measure on $(X, K)$. Since $K\mu$ is continuous, it attains its infimum and supremum on the compact set supp $\mu$. Take $x^-, x^+ \in$ supp $\mu$ such that $(K\mu)(x^-) = \inf_{\text{supp} \mu} K\mu$, $(K\mu)(x^+) = \sup_{\text{supp} \mu} K\mu$.

To prove that $\mu$ is balanced, it will suffice to show that $(K\mu)(x^-) = (K\mu)(x^+)$. Let $\varepsilon > 0$. We first construct functions $u^\pm$ such that the measures $u^\pm \mu$ approximate the Dirac measures at $x^\pm$, using Lemma 2.7. Write

$$E = \{(K\mu)^{q-1} \mid \text{supp} \mu \} \cup \{K(x, -) \mid \text{supp} \mu : x \in X \} \subseteq C(\text{supp} \mu)$$

(which is well-defined by Lemma 3.6 and Proposition 3.8(i)). Then $E$ is compact, since it is the union of a singleton with the image of the compact space $X$ under the composite of continuous maps

$$X \xrightarrow{K} C(X) \xrightarrow{\text{restriction}} C(\text{supp} \mu)$$

(using Lemma 3.5). So by Lemma 2.7, we can choose a nonnegative function $u^- \in C(X)$ such that $\int_X u^- d\mu = 1$ and

$$\left| \int_X (K\mu)^{q-1} d(u^- \mu) - (K\mu)(x^-)^{q-1} \right| \leq \varepsilon,$$

$$\left| \int_X K(x, -) d(u^- \mu) - K(x, x^-) \right| \leq \varepsilon,$$

the latter for all $x \in X$. Choose $u^+$ similarly for $x^+$.

Since $u^- - u^+$ is bounded, we can choose an open interval $I \subseteq \mathbb{R}$, containing 0, such that the function $1 + t(u^- - u^+) \in C(X)$ is strictly positive for each $t \in I$. Then for each $t \in I$, we have a probability measure

$$\mu_t = (1 + t(u^- - u^+))\mu$$

25
on $X$, with $\text{supp } \mu_t = \text{supp } \mu$. Note that $\mu_0 = \mu$.

We will exploit the fact that $D^K_q(\mu_t)$ has a local maximum at $t = 0$, showing that the function $t \mapsto D^K_q(\mu_t)^{1-q}$ is differentiable at 0 and, therefore, has derivative 0 there. For each $t \in I$,

$$D^K_q(\mu_t)^{1-q} = \int (K\mu_t)^{q-1} \, d\mu + t \int (K\mu_t)^{q-1} d((u^- - u^+)\mu)$$

$$= a(t) + b(t),$$

say. (Since $\text{supp } (K\mu_t) \supseteq \text{supp } (\mu_t) = \text{supp } \mu$, the integrand $(K\mu_t)^{q-1}$ is well-defined and continuous on $\text{supp } \mu$, and both integrals are finite.) We now show that $a(t)$ and $b(t)$ are differentiable at $t = 0$, compute their derivatives there, and bound the derivatives below.

To differentiate the integral $a(t)$, we use Lemma 2.12. Choose a bounded open subinterval $J$ of $I$, also containing 0, whose closure $\overline{J}$ is a subset of $I$. We verify that the function $f : X \times J \to \mathbb{R}$ defined by

$$f(x, t) = (K\mu_t)(x)^{q-1} = \left[ (K\mu_t)(x) + tK((u^- - u^+)\mu)(x) \right]^{q-1}$$

satisfies the conditions of Lemma 2.12.

For condition 2.12(i), we have already seen that $f(-, t) = (K\mu_t)^{q-1}$ is $\mu$-integrable for each $t \in I$. For condition 2.12(ii): for all $x \in \text{supp } \mu$, the function $f(x, -)$ is differentiable on $I$ (and in particular on $J$), with derivative

$$t \mapsto \frac{\partial f}{\partial t}(x, t) = (q-1) \left[ (K\mu_t)(x) + tK((u^- - u^+)\mu)(x) \right]^{q-2} \cdot K((u^- - u^+)\mu)(x).$$

For condition 2.12(iii), this formula shows that $\partial f / \partial t$ is continuous on $(\text{supp } \mu) \times I$. Hence $|\partial f / \partial t|$ is continuous on the compact space $(\text{supp } \mu) \times \overline{J}$, and therefore bounded on $(\text{supp } \mu) \times J$, with supremum $H$, say. Let $h : X \to \mathbb{R}$ be the function with constant value $H$. Then $h$ is $\mu$-integrable and $|\partial f / \partial t(x, t)| \leq h(x)$ for all $x \in \text{supp } \mu$ and $t \in J$, as required.

We can therefore apply Lemma 2.12, which implies that $a(t)$ is differentiable at $t = 0$ and

$$a'(0) = (q-1) \int (K\mu_t)(x)^{q-2} K((u^- - u^+)\mu)(x) \, d\mu(x)$$

$$= (q-1) \int (K\mu_t)(x)^{q-2} \left( \int K(x, y) \, d((u^- - u^+)\mu)(y) \right) \, d\mu(x)$$

$$\geq (q-1) \int (K\mu_t)^{q-2} (K(-, x^-) - K(-, x^+) + 2\varepsilon) \, d\mu,$$

(5)

where the inequality follows from the defining properties of $u^-$ and $u^+$ and the fact that $q < 1$.

Now consider $b(t)$. By definition of derivative, $b$ is differentiable at 0 if and only if the limit

$$\lim_{t \to 0} \int (K\mu_t)^{q-1} d((u^- - u^+)\mu)$$

26
exists, and in that case $b'(0)$ is that limit. As $t \to 0$,

$$K\mu_t = K\mu + tK((u^- - u^+)\mu) \to K\mu$$

in $C(\text{supp } \mu)$, so $(K\mu_t)^q \to (K\mu)^q$ in $C(\text{supp } \mu)$ (by Lemma 2.3), so

$$\int_{\text{supp } \mu} (K\mu_t)^q \, (u^--u^+) \mu \to \int_{\text{supp } \mu} (K\mu)^q \, (u^--u^+) \mu.$$ 

Hence $b'(0)$ exists and is given by

$$b'(0) = \int_X (K\mu)^q \, (u^--u^+) \mu.$$ 

By the defining properties of $u^-$ and $u^+$, it follows that

$$b'(0) \geq (K\mu)(x^-)^q - (K\mu)(x^+)^q - 2\varepsilon. \quad (6)$$

Returning to equation (4), we have now shown that both $a(t)$ and $b(t)$ are differentiable at $t = 0$. So too, therefore, is $D_q^K(\mu_0)^{1-q}$. But by the maximality of $\mu$, its derivative there is 0. Hence the bounds (5) and (6) give

$$0 \geq (q - 1) \int (K\mu)^q \, (K(-, x^-) - K(-, x^+) + 2\varepsilon) \, d\mu + (K\mu)(x^-)^q - (K\mu)(x^+)^q - 2\varepsilon$$

$$= (q - 1) \left( \int (K\mu)^q \, K(x^-, -) \, d\mu - \int (K\mu)^q \, K(x^+, -) \, d\mu + 2\varepsilon \int (K\mu)^q \, d\mu \right)$$

$$+ (K\mu)(x^-)^q - (K\mu)(x^+)^q - 2\varepsilon, \quad (7)$$

using the symmetry of $K$. Consider the first integral in (7). We have $(K\mu)(x) \geq (K\mu)(x^-)$ for all $x \in \text{supp } \mu$, by definition of $x^-$. Since $q - 2 < 0$, this implies that

$$\int (K\mu)^q \, K(x^-, -) \, d\mu \leq (K\mu)(x^-)^q \int K(x^-, -) \, d\mu = (K\mu)(x^-)^q.$$ 

A similar statement holds for $x^+$. Since $q - 1 < 0$, it follows from (7) that

$$0 \geq q ((K\mu)(x^-)^q - (K\mu)(x^+)^q) - 2\varepsilon \left( (1 - q) \int (K\mu)^q \, d\mu + 1 \right). \quad (8)$$

Put $c = (1 - q) \int (K\mu)^q \, d\mu + 1 \in \mathbb{R}$. Then by (8), the defining properties of $x^-$ and $x^+$, and the fact that $0 < q < 1$,

$$2\varepsilon c \geq q ((K\mu)(x^-)^q - (K\mu)(x^+)^q) \geq 0.$$ 

Taking $\varepsilon \to 0$, we see that $(K\mu)(x^-) = (K\mu)(x^+)$, which proves the result. \qed

**Corollary 6.4** Assume that $X$ is nonempty. For each $q \in (0, 1)$, there exists a balanced $q$-maximising probability measure on $X$. 

27
Proof. Fix \( q \in (0, 1) \). The function \( D^K_q \) is continuous on the nonempty compact space \( P(X) \) (Proposition 4.8), so attains a maximum at some \( \mu \in P(X) \). By Proposition 6.3, \( \mu \) is balanced.

Thus, balanced \( q \)-maximising measures exist for arbitrarily small \( q > 0 \). In the next section, we will use a limiting argument to find a balanced 0-maximising measure. The following lemma shows that any such measure maximises diversity of all orders simultaneously.

**Lemma 6.5** For \( 0 \leq q' \leq q \leq \infty \), any balanced probability measure that is \( q' \)-maximising is also \( q \)-maximising. In particular, any balanced measure that is 0-maximising is maximising.

**Proof.** Let \( 0 \leq q' \leq q \leq \infty \) and let \( \mu \) be a balanced \( q' \)-maximising measure. Then for all \( \nu \in P(X) \),

\[
D^K_q(\nu) \leq D^K_{q'}(\nu) \leq D^K_{q'}(\mu) = D^K_q(\mu)
\]

where the first inequality follows from Proposition 4.6(ii), the second inequality from maximality of \( D^K_{q'}(\mu) \), and the equality from Lemma 5.6 and \( \mu \) being balanced. \( \square \)

For the limiting argument, we will use:

**Lemma 6.6**

\( i. \) The set of balanced probability measures is closed in \( P(X) \).

\( ii. \) For each \( q \in (0, \infty) \), the set of \( q \)-maximising probability measures is closed in \( P(X) \).

**Proof.** For (i), by Lemma 5.6 and Proposition 4.6(ii), the set of balanced measures is

\[
\{ \mu \in P(X) : D^K_1(\mu) = D^K_2(\mu) \}.
\]

But \( D^K_1, D^K_2 : P(X) \to \mathbb{R} \) are continuous (by Proposition 4.8), so by a standard topological argument, this set is closed.

Part (ii) is immediate from the continuity of \( D^K_q \). \( \square \)

### 7 The maximisation theorem

We now come to our main theorem:

**Theorem 7.1** Let \( (X, K) \) be a nonempty symmetric space with similarities.

\( i. \) There exists a probability measure \( \mu \) on \( X \) that maximises \( D^K_q(\mu) \) for all \( q \in [0, \infty] \) simultaneously.

\( ii. \) The maximum diversity \( \sup_{\mu \in P(X)} D^K_q(\mu) \) is independent of \( q \in [0, \infty] \).
Proof. We have already shown that for each \( q \in (0, 1) \) there exists a balanced \( q \)-maximising probability measure on \( X \) (Corollary 6.4). Since \( P(X) \) is compact, we can choose some \( \mu \in P(X) \) such that for every \( q > 0 \) and neighbourhood \( U \) of \( \mu \), there exist \( q' \in (0, q) \) and a balanced \( q' \)-maximising measure in \( U \). By Lemma 6.5, for every \( q > 0 \), every neighbourhood of \( \mu \) contains a balanced \( q \)-maximising measure. To prove both parts of the theorem, it suffices to show that \( \mu \) is balanced and maximising.

Since the set of balanced measures is closed (Lemma 6.6(i)), \( \mu \) is balanced. Since the set of \( q \)-maximising measures is closed for each \( q > 0 \) (Lemma 6.6(ii)), \( \mu \) is \( q \)-maximising for each \( q > 0 \). Now given any \( \nu \in P(X) \), we have \( D^K_q(\mu) \geq D^K_q(\nu) \) for all \( q > 0 \); then passing to the limit as \( q \to 0^+ \) and using the continuity of diversity in its order (Proposition 4.6(i)) gives \( D^K_0(\mu) \geq D^K_0(\nu) \). Hence \( \mu \) is 0-maximising. But \( \mu \) is also balanced, so by Lemma 6.5, \( \mu \) is maximising. \( \square \)

The symmetry hypothesis in the theorem cannot be dropped, even in the finite case ([14], Section 8).

Part (ii) of the theorem shows that maximum diversity is an unambiguous real invariant of a space, not depending on a choice of parameter \( q \):

**Definition 7.2** Let \((X, K)\) be a nonempty symmetric space with similarities. The **maximum diversity** of \((X, K)\) is

\[
D_{\text{max}}(X, K) = \sup_{\mu \in P(X)} D^K_q(\mu) \in (0, \infty),
\]

for any \( q \in [0, \infty] \). Similarly, the **maximum entropy** of \((X, K)\) is

\[
H_{\text{max}}(X, K) = \log D_{\text{max}}(X, K) = \sup_{\mu \in P(X)} H^K_q(\mu),
\]

for any \( q \in [0, \infty] \). We often abbreviate \( D_{\text{max}}(X, K) \) to \( D_{\text{max}}(X) \).

The well-definedness of maximum diversity can be understood as follows. As explained and proved in Section 6, for a maximising measure \( \mu \), all points in \( \text{supp} \mu \) are equally typical. Diversity is mean atypicality, and although the notion of mean varies with the order \( q \), all means have the property that the mean of a constant function is that constant (Remark 4.7). Thus, our maximising measure \( \mu \) has the same diversity of all orders. That diversity is \( D_{\text{max}}(X) \).

A corollary of Theorem 7.1 is that to find a measure that maximises diversity of all positive orders, it suffices to find one that maximises diversity of just one positive order.

**Corollary 7.3** Let \((X, K)\) be a symmetric space with similarities. Suppose that \( \mu \in P(X) \) is \( q \)-maximising for some \( q \in (0, \infty) \). Then \( \mu \) is maximising.

**Proof.** Fix \( q \in (0, \infty) \) and let \( \mu \) be a \( q \)-maximising measure. Then

\[
D^K_q(\mu) \leq D^K_0(\mu) \leq D_{\text{max}}(X) = D^K_0(\mu),
\]

29
so equality holds throughout. As $D^K_0(\mu) = D^K_0(\mu)$ with $q \neq 0$, Proposition 4.6(ii) implies that $\mu$ is balanced. But also $D^K_0(\mu) = D_{\text{max}}(X)$, so $\mu$ is 0-maximising. Lemma 6.5 then implies that $\mu$ is maximising.

The exclusion of the case $q = 0$ here is necessary: not every 0-maximising measure is maximising, even in the finite case ([14], end of Section 6)

Theorem 7.1 asserts the mere existence of a maximising measure and the well-definedness of maximum diversity. But we can describe the maximum diversity and maximising measures somewhat explicitly, in terms of magnitude and weight measures:

**Corollary 7.4** Let $(X, K)$ be a nonempty symmetric space with similarities.

1. We have
   
   \[ D_{\text{max}}(X) = \sup_Y |Y|, \quad (9) \]
   
   where the supremum is over the nonempty closed subsets $Y$ of $X$ admitting a positive weight measure.

2. A probability measure $\mu$ on $X$ is maximising if and only if it is equal to $\hat{\nu}$ for some positive weight measure $\nu$ on some subset $Y$ attaining the supremum in (9). In that case, $D_{\text{max}}(X) = |\text{supp} \mu|$.

**Proof.** For any $q \in [0, \infty]$,

\[ D_{\text{max}}(X) = \sup \{ D^K_q(\mu) : \mu \in P(X), \mu \text{ is balanced} \} \]

\[ = \sup \{ |Y| : \text{nonempty closed } Y \subseteq X \text{ admitting a positive weight measure} \}, \]

where (10) follows from the existence of a balanced maximising measure and (11) from Lemma 5.6. This proves (i). Every maximising measure is balanced, so (ii) also follows, again using Lemma 5.6.

When $X$ is finite, this result provides an algorithm for computing maximum diversity and maximising measures ([14], Section 7).

As an immediate consequence of Corollary 7.4(i), maximum diversity is monotone with respect to inclusion:

**Corollary 7.5** Let $X$ be a symmetric space with similarities, and let $Y \subseteq X$ be a nonempty closed subset. Then $D_{\text{max}}(Y) \leq D_{\text{max}}(X)$.

Maximum diversity is also monotone in another sense: reducing the similarity between points increases the maximum diversity. For metric spaces, this means that as distances increase, so does maximum diversity.

**Proposition 7.6** Let $X$ be a nonempty compact Hausdorff space. Let $K, K'$ be symmetric similarity kernels on $X$ such that $K(x, y) \geq K'(x, y)$ for all $x, y \in X$. Then

\[ D_{\text{max}}(X, K) \leq D_{\text{max}}(X, K'). \]
Proof. Fix $q \in [0, \infty]$. By definition of maximum diversity, it is equivalent to show that
\[
\sup_{\mu \in P(X)} D^K_q(\mu) \leq \sup_{\mu \in P(X)} D^{K'}_q(\mu).
\]
Recall that $D^K_q(\mu) = M_{q-1}(\mu, 1/K\mu)$ for all $\mu \in P(X)$. The hypotheses imply that $K\mu \geq K'\mu$ pointwise, and the power mean is increasing in its second variable, so the result follows. \qed

8 Metric spaces

The rest of this paper has a more geometric focus. We specialise to the case of a compact metric space $X = (X, d)$, using the similarity kernel $K(x, y) = e^{-d(x, y)}$ and writing $D^K_q$ as $D_X^q$.

We have seen that maximum diversity is closely related to magnitude. Here, we review some of the geometric properties of magnitude (surveyed at greater length in [15]) and their consequences for maximum diversity. We then compute the maximum diversity of several classes of metric space.

Most of the theory of the magnitude of metric spaces assumes that the space is positive definite, meaning that for every finite subset $\{x_1, \ldots, x_n\}$, the matrix $(e^{-d(x_i, x_j)})$ is positive definite. Many of the most familiar metric spaces have this property, including all subsets of $\mathbb{R}^n$ with the Euclidean or $\ell^1$ (taxicab) metric, all subsets of hyperbolic space, and all ultrametric spaces ([18], Theorem 3.6).

There are several equivalent definitions of the magnitude of a positive definite compact metric space (as shown by Meckes in [17] and [15]). The simplest is this:
\[
|X| = \sup\{|Y| : \text{finite } Y \subseteq X\}.
\]
When $X$ admits a weight measure (and in particular, when $X$ is finite), this is equivalent to Definition 5.4. Indeed, Meckes proved ([18], Theorems 2.3 and 2.4):

**Theorem 8.1** (Meckes) Let $X$ be a positive definite compact metric space. Then
\[
|X| = \sup_{\mu} \frac{\mu(X)^2}{\int_X \int_X e^{-d(x,y)} \, d\mu(x) \, d\mu(y)},
\]
where the supremum is over all $\mu \in M(X)$ such that the denominator is nonzero. The supremum is attained by $\mu$ if and only if $\mu$ is a scalar multiple of a weight measure, and if $\mu$ is a weight measure then $|X| = \mu(X)$.

(Note that the supremum is over signed measures, unlike the similar expression for maximum diversity in Example 4.5.)

Meckes also showed that weight measures are unique when they exist ([17], after Definition 3.3). However, work such as [1] has established that even for some of the most straightforward spaces (including Euclidean balls), no weight measure exists.
An important property of magnitude for positive definite spaces, immediate from the definition, is that if \( Y \subseteq X \) then \( |Y| \leq |X| \). From Corollary 7.4(i), it follows that

\[
D_{\text{max}}(X) \leq |X| \tag{12}
\]

for all positive definite compact metric spaces \( X \neq \emptyset \). Any one-point subset of \( X \) has a positive weight measure and magnitude 1, so also

\[
D_{\text{max}}(X) \geq 1.
\]

If \( X \) does not admit a weight measure then it follows from Corollary 7.4(ii) that no maximising measure on \( X \) has full support. Indeed, the apparent rarity of spaces admitting a weight measure suggests that the supremum in Corollary 7.4 runs over a rather small class of subsets \( Y \).

Corollary 7.4 implies that the problem of computing maximum diversity is closely related to the problem of computing magnitude. There are a few spaces of geometric interest whose magnitude is known exactly, including spheres with the geodesic metric (Theorem 7 of [28]), Euclidean balls of odd dimension (whose magnitude is a rational function of the radius [1, 29, 27]), and convex bodies in \( \mathbb{R}^n \) with the \( \ell_1 \) metric (Theorem 5.4.6 of [15]; the magnitude is closely related to the intrinsic volumes). But for many very simple spaces, including even the 2-dimensional Euclidean disc, the magnitude remains unknown.

On the other hand, maximum diversity is sometimes more tractable than magnitude. Meckes showed that at least for compact subsets \( X \) of \( \mathbb{R}^n \), maximum diversity is a quantity that is already known, if little explored, in potential theory: up to a known constant factor, \( D_{\text{max}}(X) \) is the Bessel capacity of order \((n + 1)/2\) of \( X \) ([17], Section 6).

In the rest of this section, we analyse the few classes of metric space for which we are able to calculate the maximum diversity exactly. In principle this includes all finite spaces, since Corollary 7.4 then provides an algorithm for calculating the maximum diversity (described in Section 7 of [14]). This class aside, all our examples are instances of the following result.

**Lemma 8.2** Let \( X \) be a nonempty positive definite compact metric space admitting a positive weight measure \( \mu \). Then:

\begin{enumerate}
  \item the normalisation \( \hat{\mu} \) of \( \mu \) is the unique maximising measure on \( X \);
  \item \( D_{\text{max}}(X) = |X| \).
\end{enumerate}

**Proof.** For (ii), since \( X \) admits a positive weight measure, Corollary 7.4(i) gives \( D_{\text{max}}(X) \geq |X| \). But the opposite inequality (12) also holds, giving \( D_{\text{max}}(X) = |X| \). Now by Corollary 7.4(ii), it follows that that \( \hat{\mu} \) is a maximising measure. For uniqueness, let \( \nu \) be any maximising measure on \( X \). Then

\[
\frac{\nu(X)}{\int_X \int_X e^{-d(x,y)} \, d\nu(x) \, d\nu(y)} = D^X_2(\nu) = D_{\text{max}}(X) = |X|,
\]

so Theorem 8.1 implies that \( \nu \) is a scalar multiple of \( \hat{\mu} \). But both are probability measures, so \( \nu = \hat{\mu} \). \( \blacksquare \)

32
Example 8.3 Let $X$ be a finite metric space with $n$ points, satisfying $d(x,y) > \log(n-1)$ whenever $x \neq y$. Then $X$ is positive definite and its unique weight measure is positive (Proposition 2.4.17 of [12]), so Lemma 8.2 applies.

Example 8.4 As shown in Theorem 2 of [28], a line segment $[0,\ell] \subset \mathbb{R}$ has weight measure

$$\frac{1}{2}(\delta_0 + \delta_\ell + \lambda_{[0,\ell]}),$$

where $\delta_x$ denotes the Dirac measure at a point $x$ and $\lambda_{[0,\ell]}$ is Lebesgue measure on $\mathbb{R}$ restricted to $[0,\ell]$. Hence

$$|[0,\ell]| = 1 + \frac{1}{2}\ell.$$

By Lemma 8.2, the maximum diversity of $[0,\ell]$ is equal to its magnitude, and its unique maximising measure is

$$\frac{\delta_0 + \delta_\ell + \lambda_{[0,\ell]}}{2 + \ell}.$$

In fact, every compact subset of $\mathbb{R}$ has a positive weight measure (by Lemma 2.8 and Corollary 2.10 of [18]), as well as being positive definite, so again, Lemma 8.2 applies.

Example 8.5 Let $X$ be a nonempty compact metric space that is homogeneous, that is, its isometry group acts transitively on points. There is a unique isometry-invariant probability measure on $X$, the Haar probability measure $\mu$ (Theorems 4.11 and 5.3 of [25]). As observed in [28] (Theorem 1), the measure

$$\frac{\mu}{\int_X e^{-d(x,y)} \, d\mu(x)}$$

is independent of $y \in X$ and is a positive weight measure on $X$. Hence

$$|X| = \frac{1}{\int_X e^{-d(x,y)} \, d\mu(x)}$$

for all $y \in X$. This is the reciprocal of the expected similarity between a random pair of points. Assuming also that $X$ is positive definite, Lemma 8.2 implies that $D_{\text{max}}(X)$ is equal to $|X|$ and the Haar probability measure $\mu$ is the unique maximising measure.

We have shown that every symmetric space with similarities has at least one maximising measure. Although some spaces have multiple maximising measures ([14], Section 9), we now show that for many metric spaces, the maximising measure is unique.

For any metric space $X$, define a symmetric bilinear form $\langle -,- \rangle_X$ on $M(X)$ by

$$\langle \mu, \nu \rangle_X = \int_X \int_X e^{-d(x,y)} \, d\mu(x) \, d\nu(y)$$

($\mu, \nu \in M(X)$).
Lemma 8.6 Let $X$ be a nonempty compact metric space such that the bilinear form $\langle -, - \rangle_X$ on $M(X)$ is positive definite. Then $X$ admits exactly one maximising measure.

Proof. Since $\langle -, - \rangle_X$ is an inner product, the function $\mu \mapsto \langle \mu, \mu \rangle_X$ on $M(X)$ is strictly convex. Its restriction to the convex subset $P(X) \subseteq M(X)$ therefore attains a minimum at most once. But for $\mu \in P(X)$,

$$D^2_X(\mu) = 1/\langle \mu, \mu \rangle_X$$

(Example 4.5), so $\mu$ minimises $\langle -, - \rangle_X$ on $P(X)$ if and only if it is 2-maximising, or equivalently maximising (by Corollary 7.3). The result follows.

We deduce that for two important classes of metric spaces, maximising measures are unique.

Proposition 8.7 Every nonempty positive definite finite metric space has exactly one maximising measure.

Proof. This is immediate from Lemma 8.6.

The following more substantial result is due to Mark Meckes (personal communication, 2019).

Proposition 8.8 (Meckes) Every nonempty compact subset of Euclidean space has exactly one maximising measure.

Proof. Let $X$ be a nonempty compact subset of $\mathbb{R}^n$. Then $X$ is positive definite, so by Lemma 2.2 of [18], $\langle \mu, \mu \rangle_X \geq 0$ for all $\mu \in M(X)$. By Lemma 8.6, it now suffices to prove that if $\langle \mu, \mu \rangle_X = 0$ then $\mu = 0$.

Let $F$ be the function on $\mathbb{R}^n$ defined by $F(x) = e^{-\|x\|^2}$. Then

$$\langle \mu, \nu \rangle_X = \int_{\mathbb{R}^n} (F * \mu) d\nu$$

($\mu, \nu \in M(X)$), where $*$ denotes convolution. By the standard properties of the Fourier transform $\hat{\cdot}$, it follows that

$$\langle \mu, \mu \rangle_X = \int_{\mathbb{R}^n} \hat{F} |\hat{\mu}|^2.$$

But $\hat{F}$ is everywhere strictly positive (Theorem 1.14 of [24]), so if $\langle \mu, \mu \rangle_X = 0$ then $\hat{\mu} = 0$ almost everywhere, which in turn implies that $\mu = 0$ (paragraph 1.7.3(b) of [23]).

9 The uniform measure

For many of the spaces $X$ that arise most often in mathematics, there is a choice of probability measure on $X$ that seems obvious or natural. For finite sets, it is
the uniform measure. For homogeneous spaces, it is Haar measure. For subsets of \( \mathbb{R}^n \) with finite nonzero volume, it is normalised Lebesgue measure. In this section, we propose a method for assigning a canonical probability measure to any compact metric space \( X \). We will call it the uniform measure.

The idea behind this method has two parts. The first is very standard in statistics: take the probability distribution that maximises entropy. For example, in the context of differential entropy of probability distributions on \( \mathbb{R} \), the maximum entropy distribution supported on a prescribed bounded interval is the uniform distribution there, and the maximum entropy distribution with a prescribed mean and variance is the normal distribution.

However, given a compact metric space \( X \), simply taking the maximising measure on \( X \) does not give a suitable notion of uniform measure in the sense above (even putting aside the question of uniqueness). The problem is the failure of scale-invariance. For many uses of metric spaces, the choice of scale factor is somewhat arbitrary: if we multiplied all the distances by a constant \( t > 0 \), we would regard the space as essentially unchanged. (Formally, scaling by \( t \) defines an automorphism of the category of metric spaces, for any of the standard notions of map between metric spaces.) But the maximising measure depends critically on the scale factor, as almost every example in the previous section shows.

There now enters the second part of the idea: pass to the large-scale limit. Thus, we define the uniform measure on a space to be the limit of the maximising measures as the scale factor increases to \( \infty \). We now make this precise.

**Definition 9.1** Let \( X = (X, d) \) be a metric space and \( t \in (0, \infty) \). We write \( td \) for the metric on \( X \) defined by \((td)(x, y) = t \cdot d(x, y)\), and \( K^t \) for the similarity kernel on \( X \) defined by \( K^t(x, y) = e^{-td(x, y)} \). We denote by \( tX \) the set \( X \) equipped with the metric \( td \).

If \( X \) is a subspace of \( \mathbb{R}^n \) then \( tX = (X, td) \) is isometric to \( \{tx : x \in X\}, d) \), where \( d \) is Euclidean distance. But for our purposes, it is better to regard the set \( X \) as fixed and the metric as varying with \( t \).

An immediate consequence of Proposition 7.6 is that the maximum diversity of a compact metric space increases monotonically with the scale factor: \( D_{\max}(tX) \) is increasing in \( t \in (0, \infty) \).

**Definition 9.2** Let \( X \) be a compact metric space. Suppose that \( tX \) has a unique maximising measure \( \mu_t \) for all \( t \gg 0 \), and that \( \lim_{t \to \infty} \mu_t \) exists in \( P(X) \). Then the uniform measure on \( X \) is \( \mu_X = \lim_{t \to \infty} \mu_t \).

This definition has the desired property of scale-invariance:

**Lemma 9.3** Let \( X \) be a compact metric space and \( t > 0 \). Then the uniform measures on \( X \) and \( tX \) are equal: \( \mu_X = \mu_{tX} \), with one side of the equality defined if and only if the other is.

**Proof.** This is immediate from the definition. \( \square \)
The next few results show that the definition of uniform measure succeeds in capturing the ‘canonical’ measure on several significant classes of space.

**Proposition 9.4** On a nonempty finite metric space, the uniform measure exists and is equal to the uniform probability measure in the standard sense.

*Proof.* Let \( X = \{x_1, \ldots, x_n\} \) be a finite metric space. For \( t > 0 \), write \( Z^t \) for the \( n \times n \) matrix with entries \( e^{-td(x_i, x_j)} \). For \( t \gg 0 \), the space \( tX \) is positive definite with positive weight measure, by Example 8.3. Expressed as a vector, the weight measure on \( tX \) (for \( t \gg 0 \)) is

\[
(Z^t)^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.
\]

The normalisation of this weight measure is the unique maximising measure \( \mu_t \) on \( tX \), by Lemma 8.2. As \( t \to \infty \), we have \( Z^t \to I \) in the topological group \( \text{GL}_n(\mathbb{R}) \), giving \( (Z^t)^{-1} \to I \) and so \( \mu_t \to (1/n, \ldots, 1/n) \).

This result shows that the uniform measure is not in general uniformly distributed; that is, balls of the same radius may have different measures.

Our concept of uniform measure also behaves well on homogeneous spaces. We restrict to those spaces \( X \) such that \( tX \) is positive definite for every \( t > 0 \), which is equivalent to the classical condition that \( X \) is of **negative type**. (For our purposes, this can be taken as the definition of negative type, but for a proof, see Theorem 3.3 of [18].)

**Proposition 9.5** On a nonempty, homogeneous, compact metric space of negative type, the uniform measure exists and is equal to the Haar probability measure.

*Proof.* Let \( X \) be such a space. The Haar probability measure \( \mu \) on \( X \) is the unique isometry-invariant probability measure on \( X \), so it is also the Haar probability measure on \( tX \) for every \( t > 0 \). Hence by Example 8.5, \( \mu_t = \mu \) for all \( t \), and the result follows trivially.

Finally, when applied to a real interval, our definition of uniform measure produces the uniform measure in the standard sense.

**Proposition 9.6** On the line segment \( [0, \ell] \) of length \( \ell > 0 \), the uniform measure exists and is Lebesgue measure restricted to \( [0, \ell] \) and normalised to a probability measure.

*Proof.* Write \( X = [0, \ell] \) and \( d \) for the metric on \( \mathbb{R} \). For each \( t > 0 \), the metric space \( tX = (X, td) \) is isometric to the interval \( [0, t\ell] \) with metric \( d \), which by Example 8.4 has unique maximising measure

\[
\frac{\delta_0 + \delta_{t\ell} + \lambda_{[0,t\ell]}}{2 + t\ell}.
\]
Transferring this measure across the isometry, $tX$ therefore has unique maximising measure

$$
\mu_t = \frac{\delta_0 + \delta_\ell + t\lambda_{[0,\ell]}}{2 + t\ell}.
$$

Hence $\mu_t \to \lambda_{[0,\ell]} / \ell$ as $t \to \infty$, as required. \hfill \Box

We claim that Proposition 9.6 extends to Euclidean spaces of arbitrary dimension, but defer the proof to a future article. Precisely, let $X$ be a compact subspace of $\mathbb{R}^n$ with nonzero volume, write $\lambda_X$ for $n$-dimensional Lebesgue measure restricted to $X$, and write $\hat{\lambda}_X$ for its normalisation to a probability measure. Our claim is that $\hat{\lambda}_X$ is the uniform measure on $X$. Since $X$ does not in general admit a weight measure, we cannot use the same strategy as in the last few propositions, and matters are more complicated.

Here, we just prove a result that is similar in spirit. We show that at large scales, $\hat{\lambda}_X$ comes arbitrarily close to maximising diversity, in the sense of the last part of the following theorem.

**Theorem 9.7** Let $X$ be a compact subspace of $\mathbb{R}^n$ with nonzero volume $\lambda(X)$. Then

$$
\lim_{t \to \infty} \frac{D_{\max}(tX)}{|tX|} = 1 \quad \text{and} \quad \lim_{t \to \infty} \frac{D_{\max}(tX)}{t^n} = \frac{\lambda(X)}{n! \omega_n},
$$

where $\omega_n$ is the volume of the unit ball in $\mathbb{R}^n$. Moreover, for all $q \in [0, \infty]$,

$$
\lim_{t \to \infty} \frac{D_q(tX)(\hat{\lambda}_X)}{D_{\max}(tX)} = 1.
$$

**Proof.** We first show that for all $t > 0$ and $q \in [0, \infty]$,

$$
|tX| \geq D_{\max}(tX) \geq D_q(tX)(\hat{\lambda}_X) \geq \frac{\lambda(X)t^n}{n! \omega_n}. \quad (13)
$$

The first inequality in (13) is an instance of (12), since $\mathbb{R}^n$ is positive definite. The second holds by definition of maximum diversity. For the third, diversity is decreasing in its order (Proposition 4.6(ii)), so it suffices to prove the case $q = \infty$. The inequality then states that

$$
\sup_{x \in X} (K^t \hat{\lambda}_X)(x) \geq \frac{\lambda(X)t^n}{n! \omega_n},
$$

or equivalently, for all $x \in X$,

$$
(K^t \hat{\lambda}_X)(x) \leq \frac{n! \omega_n}{\lambda(X)t^n}. \quad (14)
$$

Now for all $x \in X$,

$$
(K^t \hat{\lambda}_X)(x) = \int_X K^t(x, y) \, d\hat{\lambda}_X(y) \leq \frac{1}{\lambda(X)} \int_{\mathbb{R}^n} e^{-t\|x-y\|^2} \, dy.
$$
The integral is \( n! \omega_n / t^n \), by a standard calculation (as in Lemma 3.5.9 of [12]). So we have now proved inequality (14) and, therefore, the last of the inequalities (13).

Dividing (13) through by \( |tX| \) gives

\[
1 \geq \frac{D_{max}(tX)}{|tX|} \geq \frac{D_q(tX,\lambda X)}{|tX|} \geq \frac{\lambda(X)t^n}{n!\omega_n|tX|}
\]

for all \( t > 0 \) and \( q \in [0, \infty] \). Theorem 1 of [1] states, in part, that the final term converges to 1 as \( t \to \infty \). Hence all terms do, and the result follows.

The first part of Theorem 9.7, that \( D_{max}(X)/|tX| \to 1 \) as \( t \to \infty \), is one of a collection of results expressing the intimate relationship between maximum diversity and magnitude. Perhaps the deepest of these is a result of Meckes, which uses the description of maximum diversity as a Bessel capacity (mentioned in Section 8) to establish that for each \( n \geq 1 \), there is a constant \( \kappa_n \) such that

\[
|X| \leq \kappa_n D_{max}(X)
\]

for all nonempty compact \( X \subset \mathbb{R}^n \) (Corollary 6.2 of [17]). This is a companion to the elementary fact that \( D_{max}(X) \leq |X| \) (inequality (12)).

The second equation in Theorem 9.7 implies, in particular, that one can recover the volume of \( X \subset \mathbb{R}^n \) from the asymptotic behaviour of the function \( t \mapsto D_{max}(tX) \). This result is in the same spirit as Theorem 1.1, which states that one can also recover the Minkowski dimension of \( X \). Thus, the asymptotics of \( D_{max}(tX) \) contain fundamental geometric information.

### 10 Open questions

(1) Maximum diversity is a numerical invariant of compact metric spaces (and more generally, symmetric spaces with similarity). What properties does this invariant have with respect to products, unions, etc., of spaces? Similarly, what are the maximising measures on a product or union of spaces, and what is the uniform measure?

(2) We know almost nothing about the maximising measures on specific non-finite metric spaces. For instance, what is the maximising measure on a Euclidean ball or cube? We do not even know the support of the maximising measure. We conjecture that in the case of a Euclidean ball, the support of the maximising measure is a finite union of concentric spheres.

(3) The uniform measure, when defined, provides a canonical way of equipping a metric space with a probability measure. But so too does the Hausdorff measure. More exactly, if the Hausdorff dimension \( d \) of \( X \) is finite then we have the Hausdorff measure \( \mathcal{H}^d \) on \( X \), which if \( 0 < \mathcal{H}^d(X) < \infty \) can be normalised to a probability measure on \( X \). What is the relationship between the Hausdorff probability measure and the uniform measure? The fact that the growth of \( D_{max}(tX) \) as \( t \to \infty \) is the Minkowski rather than Hausdorff dimension suggests that the relationship is probably not simple.
(4) What is the relationship between our notion of uniform measure on a compact metric space and that proposed by Ostrovsky and Sirota [21] (which is based on entropy of a different kind)?

(5) For finite spaces with similarity, the diversity measures studied here were first introduced in ecology [13] and have been successfully applied there. What are the biological applications of our diversity measures on infinite compact spaces? It may seem implausible that there could be any, since the points of the space are usually interpreted as species. However, in microbial biology it is common to treat the space of possible organisms as a continuum. Sometimes groupings are created, such as serotypes (strains) of a virus or operational taxonomic units (genetically similar classes) of bacteria, but it is recognised that such divisions can be arbitrary. What information do our diversity measures, and maximum diversity, convey about continuous spaces of organisms?

References

[1] J. A. Barceló and A. Carbery. On the magnitudes of compact sets in Euclidean spaces. American Journal of Mathematics, 140(2):449–494, 2018.

[2] F. Borceux. Handbook of Categorical Algebra 2: Categories and Structures. Cambridge University Press, Cambridge, 1994.

[3] N. Bourbaki. Integration: Chapitres 1 à 4. Eléments de Mathématique. Springer, Berlin, 1965.

[4] N. Dunford and J. T. Schwartz. Linear Operators. Part I: General Theory. Wiley-Interscience, New York, 1958.

[5] H. Gimperlein and M. Goffeng. On the magnitude function of domains in Euclidean space. Preprint arXiv:1706.06839, available at arXiv.org, 2017.

[6] M. Gromov. Metric Structures for Riemannian and Non-Riemannian Spaces. Birkhäuser, Boston, 2001.

[7] G. H. Hardy, J. E. Littlewood, and G. Pólya. Inequalities. Cambridge University Press, Cambridge, 2nd edition, 1952.

[8] M. O. Hill. Diversity and evenness: a unifying notation and its consequences. Ecology, 54(2):427–432, 1973.

[9] A. Klenke. Probability Theory: A Comprehensive Course. Universitext. Springer, New York, 2013.

[10] F. W. Lawvere. Metric spaces, generalized logic, and closed categories. Rendiconti del Seminario Matematico e Fisico di Milano, 43:135–166, 1973.

[11] T. Leinster. The Euler characteristic of a category. Documenta Mathematica, 13:21–49, 2008.

[12] T. Leinster. The magnitude of metric spaces. Documenta Mathematica, 18:857–905, 2013.

[13] T. Leinster and C. A. Cobbold. Measuring diversity: the importance of species similarity. Ecology, 93(3):477–489, 2012.

[14] T. Leinster and M. W. Meckes. Maximizing diversity in biology and beyond. Entropy, 18(3), 2016.
[15] T. Leinster and M. W. Meckes. The magnitude of a metric space: from category theory to geometric measure theory. In N. Gigli, editor, Measure Theory in Non-Smooth Spaces. de Gruyter Open, Warsaw, 2017.

[16] A. E. Magurran. Measuring Biological Diversity. Blackwell, Oxford, 2004.

[17] M. W. Meckes. Magnitude, diversity, capacities, and dimensions of metric spaces. Potential Analysis, 42:549–572, 2015.

[18] M. W. Meckes. Positive definite metric spaces. Positivity, 17(3):733–757, 2013.

[19] J. R. Munkres. Topology. Prentice Hall, Upper Saddle River, NJ, second edition, 2000.

[20] C. P. Niculescu and L.-E. Persson. Convex Functions and Their Applications: A Contemporary Approach. Canadian Mathematical Society Books in Mathematics. Springer, New York, 2006.

[21] E. Ostrovsky and L. Sirota. Uniform measures on the arbitrary compact metric spaces. Preprint arXiv:1403.5725, available at arXiv.org, 2014.

[22] A. Rényi. On measures of entropy and information. In J. Neymann, editor, 4th Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pages 547–561. University of California Press, 1961.

[23] W. Rudin. Fourier Analysis on Groups. Interscience, New York, 1962.

[24] E. M. Stein and G. Weiss. Introduction to Fourier Analysis on Euclidean Spaces. Princeton Mathematical Series 32. Princeton University Press, Princeton, N.J., 1971.

[25] R. C. Steinlage. Haar measure in locally compact $T_2$ spaces. American Journal of Mathematics, 97:291–307, 1975.

[26] S. D. Veresoglou, J. R. Powell, J. Davison, Y. Lekberg, and M. C. Rillig. The Leinster and Cobbold indices improve inferences about microbial diversity. Fungal Ecology, 11:1–7, 2014.

[27] S. Willerton. On the magnitude of odd balls via potential functions. Preprint arXiv:1804.02174, available at arXiv.org, 2018.

[28] S. Willerton. On the magnitude of spheres, surfaces and other homogeneous spaces. Geometriae Dedicata, 168(1):291–310, 2014.

[29] S. Willerton. The magnitude of odd balls via Hankel determinants of reverse Bessel polynomials. Preprint arXiv:1708.03227, available at arXiv.org, 2017.