Author/s:
Bailey, J;Houle, ME;Ma, X

Title:
Local Intrinsic Dimensionality, Entropy and Statistical Divergences

Date:
2022-09-01

Citation:
Bailey, J., Houle, M. E. & Ma, X. (2022). Local Intrinsic Dimensionality, Entropy and Statistical Divergences. Entropy, 24 (9), https://doi.org/10.3390/e24091220.

Persistent Link:
https://hdl.handle.net/11343/322263

entropy

MDPI

Article

Local Intrinsic Dimensionality, Entropy and Statistical Divergences

James Bailey [1,*], Michael E. Houle [1,†] and Xingjun Ma [2]

[1] School of Computing and Information Systems, The University of Melbourne, Melbourne, VIC 3010, Australia
[2] School of Computer Science, Fudan University, Shanghai 200437, China
* Correspondence: baileyj@unimelb.edu.au
† This work was partially conducted when M.E.H. was with the National Institute of Informatics, Japan.

Abstract: Properties of data distributions can be assessed at both global and local scales. At a highly localized scale, a fundamental measure is the local intrinsic dimensionality (LID), which assesses growth rates of the cumulative distribution function within a restricted neighborhood and characterizes properties of the geometry of a local neighborhood. In this paper, we explore the connection of LID to other well known measures for complexity assessment and comparison, namely, entropy and statistical distances or divergences. In an asymptotic context, we develop analytical new expressions for these quantities in terms of LID. This reveals the fundamental nature of LID as a building block for characterizing and comparing data distributions, opening the door to new methods for distributional analysis at a local scale.

Keywords: entropy; tail entropy; cumulative entropy; entropy power; intrinsic dimensionality; local intrinsic dimension; statistical divergences; statistical distances

Citation: Bailey, J.; Houle, M.E.; Ma, X. Local Intrinsic Dimensionality, Entropy and Statistical Divergences. *Entropy* **2022**, *24*, 1220. https://doi.org/10.3390/e24091220

Academic Editors: Steeve Zozor, Mariela Portesi, Pedro W. Lamberti, Gustavo Martin Bosyk and Jean-François Bercher

Received: 4 July 2022
Accepted: 26 August 2022
Published: 30 August 2022

1. Introduction

Fundamental activities for analyzing data include both an ability to characterize data complexity and an ability to make comparisons between distributions. Widely used measures for these activities include entropy (for assessing uncertainty) and statistical divergences or distances (to compare distributions) [1]. Such analysis can be performed at either a global scale across the entire data distribution or at a local scale, in the vicinity of a given location in the distribution.

An important measure of global complexity is intrinsic dimensionality, which captures the effective number of degrees of freedom needed to describe the entire dataset. On the other hand, *local intrinsic dimensionality (LID)* [2] is capable of characterizing the complexity of the data distribution around a specified query location, thus capturing the number of degrees of freedom present at a local scale. LID is a unitless quantity that can also be interpreted as a relative growth rate of probability measure within an expanding neighborhood around the specified query location, or the intrinsic dimension of the space immediately around the query point.

Our focus in this paper is to characterize entropy and statistical divergences at a highly local scale, for an asymptotically small vicinity around a specified location. We show that it is possible to leverage properties that arise from LID based characterizations of lower tail distributions [3], to develop analytical expressions for a wide selection of entropy variants and statistical divergences, in both univariate and multivariate settings. This yields expressions for *tail entropies* and *tail divergences*.

Analytical characterizations for tail divergences and tail entropies are appealing from a number of perspectives. These are as follows:

- For univariate scenarios, if working with the tail of a distribution that has a single variable, we can conduct:

  - Temporal analysis: when a distribution models some property varying over time (e.g., survival analysis), we can analyze the entropy of a univariate distribution within an asymptotically short window of time, or the divergence between two univariate distributions within an asymptotically short window of time.
  - Distance-based analysis: when a distribution models distances from a query location to its nearest neighbors and the distances are induced by a global data distribution. Here, our results can be used for analysis of tail entropy or divergence between distributions within an asymptotically small distance interval. In the case of the latter, this can provide insight into multivariate properties, since under minimal assumptions the divergences between univariate distance distributions provide lower bounds for distances between multivariate distributions [4,5]. This is applicable for models such as generative adversarial networks (GANs), where it is important to test correspondence between synthetic and true distributions at a local level [6].
- For multivariate scenarios where we are analyzing distributions with multiple variables:
  - If an assumption of locally spherical symmetry of the distribution holds, then we can directly compute the tail entropy of a distribution or the divergence between two tail distributions in the vicinity of a single point. Such an assumption is suitable for analyzing data distributions for many types of physical systems such as fluids, glasses, metals and polymers, where local isotropy holds.

A key challenge in developing analytical characterizations for tail entropies and tail divergences is how to avoid or minimize assumptions about the form of the local distribution in the vicinity of the query (for example, assumptions such as a local normal distribution or a local uniform distribution). As we will see, analytical results are in fact possible—as the neighborhood radius asymptotically tends to zero, the tail distribution (a truncated distribution induced from the global distribution) is guaranteed to converge to a generalized pareto distribution (GPD), with the GPD parameter determined by the LID value of the tail distribution. The technical challenge is to rigorously delineate under what circumstances it is possible to leverage this relationship to achieve a dramatic simplification of the integrals that are required to compute varieties of tail entropy or distribution divergences. Our results in this paper show that such simplifications are in fact possible, for a wide range of tail entropies and divergences. This allows us to characterize and analyze fundamental properties of local neighborhood geometry, with results holding asymptotically for essentially all smooth data distributions.

In summary, our key contributions are the development of substantial new theory that asymptotically relates tail entropy, divergences and LID. It builds on and extends an earlier work by Bailey et al. [3], which focused solely on univariate entropies, without reference to divergences or multivariate settings. Specifically in this paper, we:

- Formulate technical lemmas which delineate when it is possible to substitute certain types of tail distributions by simple formulations that depend only on their associated LID values.
- Use these lemmas to compute univariate tail formulations of entropy, cross entropy, cumulative entropy, entropy power and generalized q-entropies, all in terms of the LID values of the original tail distributions.
- Use these lemmas to compute tail formulations of univariate statistical divergences and distances (Kullback–Leibler divergence, Jensen–Shannon divergence, Hellinger distance, χ^2 divergence, α-divergence, Wasserstein distance and L_2 distance).
- Extend the univariate results to a multivariate context, when local spherical symmetry of the distribution holds.

2. Related Work

The core of our study involves intrinsic dimensionality (ID) and we begin by reviewing previous work on this topic.

There is a long history of work on ID, and this can be assessed either globally (for every data point) or locally (with respect to a chosen query point). Surveys of the field provide a good overview [7–9]. In the global case, a range of previous works have focused on topological models and appropriate estimation methods [10–12]. Such examples encompass techniques such as PCA and its variants [13], graph based methods [14] and fractal models [7,15]. Other approaches such as IDEA [16,17], DANCo [18] or 2-NN estimate the (global) intrinsic dimension based on concentration of norms and angles, or 2-nearest neighbors [19].

Local intrinsic dimensionality focuses on the intrinsic dimension of a particular query point and has been used in a range of applications. These include modeling deformation in granular materials [20,21], climate science [22,23], dimension reduction via local PCA [24], similarity search [25], clustering [26], outlier detection [27], statistical manifold learning [28], adversarial example detection [29], adversarial nearest neighbor characterization [30,31] and deep learning understanding [32,33]. In deep learning, it has been shown that adversarial examples are associated with high LID estimates, a characteristic that can be leveraged to build accurate adversarial example detectors [29]. It has also been found that the LID of deep representations [33] learned by Deep Neural Networks (DNNs) or the raw input data [34,35] is correlated with the generalization performance of DNNs. A 'dimensionality expansion' phenomenon has been observed when DNNs overfit to noisy class labels [32] and this can be leveraged to develop improved loss functions. The use of a "cross-LID" measure to evaluate the quality of synthetic examples generated by GANs has been proposed in [36]. Connections between local intrinsic dimensionality and global intrinsic dimensionality were explored by Romano et al in [37]. In the area of climate science and dynamical systems, a formulation similar to local intrinsic dimensionality has been developed and referred to as local dimension or instantaneous dimension [22,23,38], using links to extreme value theoretic methods. It has proved useful as measure to characterize predictability of states and explain system dynamics.

For local intrinsic dimensionality, a popular estimator is the maximum likelihood estimator, studied in the Euclidean setting by Levina and Bickel [39] and later formulated under the more general assumptions of extreme value theory by Houle [2] and Amsaleg et al. [40], who showed it to be equivalent to the classic Hill estimator [41]. Other local estimators include expected simplex skewness [42], the tight locality estimator [43], the MiND framework [17], manifold adaptive dimension [44], statistical distance [45] and angle-based approaches [46]. Smoothing approaches for estimation have also been used with success [47,48].

Local intrinsic dimensionality is closely related to (univariate) distance distributions. Fundamental relations for interpoint distances, connecting multivariate distributions and univariate distributions have been explored by both [4,5]. The former showed that two multivariate distributions are equal whenever the interpoint distances both within and between samples have the same univariate distribution, while the latter showed that two multivariate distributions F and G are different if their univariate distance distributions from some randomly chosen point z are different. This can form the basis of a two sample test for comparing F and G. These studies have implications for our work in this paper, since they characterize the role that comparison between univariate distributions can play as a necessary condition for comparing equality of multivariate distributions.

Our work in this paper formulates results for different varieties of entropy and different types of divergences. Entropy is a fundamental notion used across many scientific disciplines. A good overview of its role in information theory is presented in [49]. Entropy power (the exponential of entropy) is commonly used in signal processing and information theory, and is a building block for the well-known Shannon entropy power inequality which can be used to analyze the convolution of two independent random variables [50]. Entropy

power goes under the name of perplexity in the field of natural language processing [51] and true diversity in the field of ecology [52]. It also corresponds to the volume of the smallest set that contains most of the probability measure [49], and it can be interpreted as a measure of statistical dispersion [53]. It is also related to Fisher information via Stam's inequality [54].

Cumulative entropy was formulated in [55] and is a modification of cumulative residual entropy [56]. It is popular in reliability theory where it is used to characterize uncertainty over time intervals. Apart from reliability theory analysis, it has been used in data mining tasks such as dependency analysis [57] and subspace cluster analysis [58], where it has proved more effective due to good estimation properties. These data mining investigations have used cumulative entropy at a global level (over the entire data domain), rather than at the local (tail) level, as in our study. Generalized variants based on Tsallis q-statistics have been developed for both entropy [59] and cumulative entropy [60]. Inclusion of the extra q parameter can assist with higher robustness to anomalies and better fitting to characteristics of data distributions. Tail entropy has been used in financial applications for measuring the expected shortfall [61] in the upper tail using quantization. This is different from our context, where the our exclusive focus is on lower tails and we develop exact results for an asymptotic regime where lower tail size approaches zero.

Divergences between probability distributions are a fundamental building block in statistics and are used to assess the degree to which one probability distribution is different from another probability distribution. They have a wide range of formulations [1] and applications, which range from use as objective functions in supervised and unsupervised machine learning [62], to hypothesis and two sample or goodness of fit testing in statistics [63], as well as generative modeling in deep learning, particularly using the Wasserstein distance [64]. Asymptotic forms of KL divergence have been investigated by Contreras-Reyes [65], for comparison of multivariate asymmetric heavy-tailed distributions.

Finally, we note that this work considerably expands a recent study by Bailey et al. [3], which established relationships between tail entropies and LID. This current paper extends and generalizes that work in several directions: (i) We establish general lemmas that provide sufficient conditions for when it is possible to substitute a tail distribution with components such as a power law, inside an integral. The techniques of [3] were specially crafted for specific integrals. (ii) We provide results for statistical divergences and distances (the work of [3] only considers entropy). (iii) We show how to formulate results for the multivariate context (as [3] only considers univariate scenarios).

3. Local Intrinsic Dimensionality

In this section, we summarize the LID model using the presentation of [2]. LID can be regarded as a continuous extension of the expansion dimension [66,67]. Like earlier expansion-based models of intrinsic dimension, its motivation comes from the relationship between volume and radius in an expanding ball, where (as originally stated in [68]) the volume of the ball is taken to be the probability measure associated with the region it encloses. The probability as a function of radius—denoted by $F(r)$—has the form of a univariate cumulative distribution function (CDF). The model formulation (as stated in [2]) generalizes this notion to real-valued functions F for which $F(0) = 0$, under appropriate assumptions of smoothness.

Definition 1 ([2]). *Let F be a real-valued function that is non-zero over some open interval containing $r \in \mathbb{R}$, $r \neq 0$. The* intrinsic dimensionality of F at r *is defined as follows whenever the limit exists:*

$$\mathrm{IntrDim}_F(r) \triangleq \lim_{\epsilon \to 0} \frac{\ln(F((1+\epsilon)r)/F(r))}{\ln(1+\epsilon)}.$$

When F satisfies certain smoothness conditions in the vicinity of r, its intrinsic dimensionality has a convenient known form:

Theorem 1 ([2]). *Let F be a real-valued function that is non-zero over some open interval containing $r \in \mathbb{R}$, $r \neq 0$. If F is continuously differentiable at r and using $F'(r)$ to denote the derivative $\frac{dF(r)}{dr}$, then*

$$\mathrm{ID}_F(r) \triangleq \frac{r \cdot F'(r)}{F(r)} = \mathrm{IntrDim}_F(r)\,.$$

Let $\mathbf{x}$ be a location of interest within a data domain $\mathcal{S}$ for which the distance measure $d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_+ \cup 0$ has been defined. To any generated sample $\mathbf{s} \in \mathcal{S}$ we associate the distance $d(\mathbf{x}, \mathbf{s})$; in this way, a *global* distribution that produces the sample $\mathbf{s}$ can be said to induce the random value $d(\mathbf{x}, \mathbf{s})$ from a *local* distribution of distances taken with respect to $\mathbf{x}$. The CDF $F(r)$ of the local distance distribution is simply the probability of the sample distance lying within a threshold r—that is, $F(r) \triangleq \mathsf{Pr}[d(\mathbf{x}, \mathbf{s}) \leq r]$. In characterizing the local intrinsic dimensionality in the vicinity of location $\mathbf{x}$, we are interested in the limit of $\mathrm{ID}_F(r)$ as the distance r tends to 0, which we denote by

$$\mathrm{ID}_F^* \triangleq \lim_{r \to 0} \mathrm{ID}_F(r)\,.$$

Henceforth, when we refer to the local intrinsic dimensionality (LID) of a function F, or of a point $\mathbf{x}$ whose induced distance distribution has F as its CDF, we will take 'LID' to mean the quantity ID_F^*. In general, ID_F^* is not necessarily an integer. In practice, estimation of the LID at $\mathbf{x}$ would give an indication of the dimension of the submanifold containing $\mathbf{x}$ that best fits the distribution.

The function ID_F can be seen to fully characterize its associated function F. This result is analogous to a foundational result from the statistical theory of extreme values (EVT), in that it corresponds under an inversion transformation to the Karamata representation theorem [69] for the upper tails of regularly varying functions. For more information on EVT and how the LID model relates to the extreme-value theoretic generalized pareto distribution, we refer the reader to [2,70,71].

Theorem 2 (**LID Representation Theorem** [2]). *Let $F : \mathbb{R} \to \mathbb{R}$ be a real-valued function, and assume that ID_F^* exists. Let x and w be values for which x/w and $F(x)/F(w)$ are both positive. If F is non-zero and continuously differentiable everywhere in the interval $[\min\{x, w\}, \max\{x, w\}]$, then*

$$\frac{F(x)}{F(w)} = \left(\frac{x}{w}\right)^{\mathrm{ID}_F^*} \cdot A_F(x, w), \quad \textit{where} \quad A_F(x, w) \triangleq \exp\left(\int_x^w \frac{\mathrm{ID}_F^* - \mathrm{ID}_F(t)}{t}\,\mathrm{d}t\right),$$

whenever the integral exists.

In [2], conditions on x and w are provided for which the factor $A_F(x, w)$ can be seen to tend to 1 as $x, w \to 0$. The convergence characteristics of F to its asymptotic form are expressed by the factor $A_F(x, w)$, which is related to the slowly varying component of functions as studied in EVT [70]. As we will shown in the next section, we make use of the LID Representation Theorem in our analysis of the limits of tail entropy variants under a form of normalization.

4. Definitions of Tail Entropies and Tail Dissimilarity Measures

In this section, we present the formulations of entropy, divergences and distances that will be studied in the later sections, in the light of the model of local intrinsic dimensionality outlined in Section 3. These entropies and dissimilarity measures will all be conditioned on the lower tails of smooth functions on domains bounded from below at zero. In each case, the formulations involve one or more non-negative real-valued functions whose restriction to $[0, w]$ satisfies certain smooth growth properties:

Definition 2. *Let $F : \mathbb{R}_+ \cup 0 \to \mathbb{R}_+ \cup 0$ be a function that is positive except at $F(0) = 0$. We say that F is a* smooth growth function *if*

- *There exists a value $r > 0$ such that F is monotonically increasing over $(0, r)$;*
- *F is continuous over $[0, r)$;*
- *F is differentiable over $(0, r)$; and*
- *The local intrinsic dimensionality ID_F^* exists and is positive.*

Given a smooth growth function F and a value $w > 0$, we define $F_w(t) \triangleq \frac{F(t)}{F(w)}$. If F is the CDF of some random variable $X \geq 0$, then $F_w(t) = \mathsf{Pr}[X \leq t \,|\, X \leq w]$, which can in turn be interpreted as the CDF of the distribution of X conditioned to the lower tail $[0, w]$. It is easy to see that for a sufficiently small choice of w, F_w must also be a smooth growth function. Its derivative $F'_w(t) = \frac{F'(t)}{F(w)}$ exists since $F'(t)$ exists, and thus can be regarded as the probability density function (PDF) of the restriction of F to $[0, w]$. In addition, it can easily be shown (using Theorem 1) that the LID of F_w is identical to that of F.

If the monotonicity of the function F is strict over the domain of interest $[0, r)$, its inverse function F^{-1} exists and satisfies the smooth growth conditions within some neighborhood of the origin. Moreover, F_w^{-1} is also a smooth growth function over $[0, 1]$, with $F_w^{-1}(0) = 0$ and $F_w^{-1}(1) = w$.

The following tail entropy, tail divergence and tail distance formulations all apply to any functions F and G satisfying the conditions stated above; in particular, they involve one or more of F_w, F'_w, G_w, G'_w, and (if the monotonicity of the functions is strict) F_w^{-1} and G_w^{-1}. In their definitions, the only difference between the tail variants and the original versions is that the distributions are conditioned on the lower tail $[0, w]$. In the tail measures involving one or more of F_w, F'_w, G_w and G'_w, integration is performed over the lower tail and not the entire distributional range $[0, +\infty)$; for the variant involving F_w^{-1} and G_w^{-1}, integration is performed over $[0, 1]$ for values of w constrained to the lower tail.

We begin with (differential) tail entropy. Entropy is perhaps the most fundamental and widely used model of data complexity and can be regarded as a measure of the uncertainty of a distribution. Differential entropy assesses the expected surprisal of a random variable and can take negative values.

Definition 3 (**Tail Entropy**). *The entropy of F conditioned on $[0, w]$ is*

$$\mathrm{H}(F, w) \triangleq -\int_0^w F'_w(t) \ln F'_w(t) \, \mathrm{d}t \,.$$

The tail entropy is equal to $E(-\log F'_w)$, the expected value of the (tail) log-likelihood. It is also possible to define the variance of the (tail) log-likelihood. This is known as the *varentropy*. Understanding this further, note that one may define the information content of a random variable X with density function f, to be $-\log f(X)$. The entropy (uncertainty) then corresponds to the expected value of the information content of X and the varentropy corresponds to the variance of the information content of X. The varentropy was introduced by Song [72] as an intrinsic measure of the shape of a distribution and has been explored in a range of studies [73–75].

Definition 4 (**Tail varentropy**). *The varentropy of F conditioned on $[0, w]$ is*

$$\mathrm{VarH}(F, w) \triangleq \int_0^w F'_w(t) \ln^2 F'_w(t) \, \mathrm{d}t \; - \left(\int_0^w F'_w(t) \ln F'_w(t) \, \mathrm{d}t \right)^2$$

The cumulative entropy is a variant of entropy proposed in [55,56] due to its attractive theoretical properties. Tail conditioning on the cumulative entropy has the same general form as that of the tail entropy. Cumulative entropy [55,56] is an information-theoretic measure popular in reliability theory, where it is used to model uncertainty over time intervals. It corresponds to the expected value of the mean inactivity time. Compared to

ordinary Shannon differential entropy, cumulative entropy has certain attractive properties, such as non-negativity and ease of estimation.

Definition 5 (**Cumulative Tail Entropy**). *The cumulative entropy of F conditioned on $[0, w]$ is*

$$\mathrm{cH}(F, w) \triangleq -\int_0^w F_w(t) \ln F_w(t)\, \mathrm{d}t\,.$$

The entropy power is the exponential of the entropy, and is also known as *perplexity* in the natural language processing community. It corresponds to the volume of the smallest set that contains most of the probability measure [49], and can be interpreted as a measure of statistical dispersion [53]. There are several standard definitions of entropy power in the research literature. For our purposes, we adopt the simplest—the exponential of Shannon entropy—for our definition conditioned to the tail.

Definition 6 (**Tail Entropy Power**). *The entropy power of F conditioned on $[0, w]$ is defined to be*

$$\mathrm{HP}(F, w) \triangleq \exp\left(\mathrm{H}(F, w)\right).$$

In the introduction, we briefly mentioned some motivation for the entropy power $\mathrm{HP}(F, w)$. We can add to this as follows:

- It can be interpreted as a diversity. Observe that when F is a (univariate) uniform distance distribution ranging over the interval $[0, w]$, we have $\mathrm{ID}_F^* = 1$ and $\mathrm{HP}(F, w) = w$. In other words, the entropy power is equal to the 'effective diversity' of the distribution (the number of neighbor distance possibilities).
- Given two different queries, each with its own neighborhood, one query with tail entropy power equal to 2 and the other with tail entropy power equal to 4, we can say that the distance distribution of the second query is twice as diverse as that of the first query.

For each of the tail entropy variants introduced above, we also propose analogous variants based on the q-entropy formulation due to Tsallis [59]. Generalized Tsallis entropies [59,60] are a family of entropies characterized via an exponent parameter q applied to the probabilities, in which the traditional (Shannon) entropy variants are obtained as the special case when q is allowed to tend to 1. The use of such a parameter q can often facilitate more accurate fitting of data characteristics and robustness to outliers.

Definition 7 (**Tail q-Entropy**). *For any $q > 0$ $(q \neq 1)$, the q-entropy of F conditioned on $[0, w]$ is defined to be*

$$\mathrm{H}_q(F, w) \triangleq \frac{1}{q-1}\left(1 - \int_0^w \left(F_w'(t)\right)^q \mathrm{d}t\right) = \frac{1}{q-1}\int_0^w F_w'(t) - \left(F_w'(t)\right)^q \mathrm{d}t\,.$$

Definition 8 (**Cumulative Tail q-Entropy**). *For any $q > 0$ $(q \neq 1)$, the cumulative q-entropy of F conditioned on $[0, w]$ is defined to be*

$$\mathrm{cH}_q(F, w) \triangleq \frac{1}{q-1}\int_0^w F_w(t) - \left(F_w(t)\right)^q \mathrm{d}t\,.$$

We define the tail q-entropy power using the q-exponential function from Tsallis statistics [59], $\exp_q(x) \triangleq [1 + (1-q)x]^{\frac{1}{1-q}}$. Note that L'Hôpital's rule can be used to show that $\exp_q(x) \to e^x$ as $q \to 1$.

Definition 9 (Tail *q*-Entropy Power). *For any $q > 0$ ($q \neq 1$), the q-entropy power of F conditioned on $[0, w]$ is defined to be*

$$\mathrm{HP}_q(F, w) \triangleq \left[1 + (1-q)\mathrm{H}_q(F, w)\right]^{\frac{1}{1-q}}.$$

We next define the tail cross entropy. Cross entropy can be used to compare two probability distributions and is often employed as a loss function in machine learning, comparing a true distribution and a learned distribution. From an information theoretic perspective, cross entropy corresponds to the expected coding length when a wrong distribution G is assumed while the data actually follows a distribution F.

Definition 10 (Tail Cross Entropy). *The cross entropy from F to G, conditioned on $[0, w]$, is defined to be*

$$\mathrm{XH}(F; G, w) \triangleq -\int_0^w F'_w(t) \ln G'_w(t)\, \mathrm{d}t.$$

Similar to entropy power, we can also define the cross entropy power, which is the exponential of the cross entropy.

Definition 11 (Tail Cross Entropy Power). *The cross entropy power from F to G, conditioned on $[0, w]$, is defined to be*

$$\mathrm{XHP}(F; G, w) \triangleq \exp\left(-\int_0^w F'_w(t) \ln G'_w(t)\, \mathrm{d}t\right).$$

A classic and fundamental method for comparing two probability distributions is the Kullback–Leibler divergence (KL Divergence) [76]. $KL(F, G)$ measures the degree to which a probability distribution G is different from a reference probability distribution F. It is a member of both the family of f-divergences and Bregman divergences. It is widely used in statistics, machine learning and information theory.

Definition 12 (Tail KL Divergence). *The Kullback–Leibler divergence from F to G, conditioned on $[0, w]$, is defined to be*

$$\mathrm{KL}(F; G, w) \triangleq \int_0^w F'_w(t) \ln \frac{F'_w(t)}{G'_w(t)}\, \mathrm{d}t.$$

The tail KL divergence can be connected to the tail entropy and the tail cross entropy according to the relationship $\mathrm{KL}(F; G, w) = \mathrm{XH}(F; G; w) - \mathrm{H}(F, w)$.

The Jensen–Shannon divergence (JS divergence) [77] is another popular measure of distance between probability distributions. It is based on the KL divergence, but unlike the KL, the square root of the JS divergence is a true metric.

Definition 13 (Tail JS Divergence). *The Jensen–Shannon divergence between F and G, conditioned on $[0, w]$, is defined to be*

$$\mathrm{JS}(F; G, w) \triangleq \frac{\mathrm{KL}(F; M, w) + \mathrm{KL}(G; M, w)}{2}, \text{ where } M(t) = \frac{F(t) + G(t)}{2}.$$

The tail JS divergence can also be written in terms of the tail entropies $\mathrm{JS}(F; G, w) = \mathrm{H}(\frac{F+G}{2}, w) - \frac{\mathrm{H}(F,w)+\mathrm{H}(G,w)}{2}$

The L2 distance is the squared Euclidean distance when comparing two probability distributions. It is part of the family of β divergences when setting $\beta = 2$ [78].

Definition 14 (Tail L2 Distance). *The L2 distance between F and G, conditioned on $[0, w]$, is defined to be*

$$\mathrm{L2D}(F; G, w) \triangleq \int_0^w \left(F'_w(t) - G'_w(t)\right)^2 \mathrm{d}t\,.$$

The Hellinger distance [79] is a true metric for comparing two probability distributions. The squared Hellinger distance a member of the family of f-divergences and is part of the family of α divergences when setting $\alpha = \frac{1}{2}$ [80].

Definition 15 (Tail Hellinger Distance). *The Hellinger distance between F and G, conditioned on $[0, w]$, is defined to be*

$$\mathrm{HD}(F; G, w) \triangleq \sqrt{\frac{1}{2}\int_0^w \left(\sqrt{F'_w(t)} - \sqrt{G'_w(t)}\right)^2 \mathrm{d}t}\,.$$

The χ^2 divergence between two probability distributions [81] is a member of the family of f divergences and is part of the family of α divergences when setting $\alpha = 2$ [80].

Definition 16 (Tail χ^2-Divergence). *The χ^2 divergence between F and G, conditioned on $[0, w]$, is defined to be*

$$\chi^2\mathrm{D}(F; G, w) \triangleq \int_0^w \frac{\left(F'_w(t) - G'_w(t)\right)^2}{G'_w(t)} \mathrm{d}t\,.$$

The asymmetric α-divergence [80] is another member of the family of f divergences. When $\alpha = 2$ it is proportional to the χ^2 divergence. When $\alpha = 0.5$ it is proportional to the squared Hellinger distance. When $\alpha \to 1$ it corresponds to the KL-divergence.

Definition 17 (Tail α-Divergence). *The α-divergence from F to G, conditioned on $[0, w]$, is defined to be*

$$\alpha\mathrm{D}(F; G, w) \triangleq \frac{1}{\alpha(1-\alpha)} \int_0^w \alpha F'_w(t) + (1-\alpha)G'_w(t) - \left(F'_w(t)\right)^\alpha \left(G'_w(t)\right)^{1-\alpha} \mathrm{d}t\,.$$

The Wasserstein distance between two probability distributions is also known as the Kantorovich–Rubinstein metric [82] or the earth mover's distance. It has become very popular as part of the loss function used in generative adversarial networks [83]. In the univariate case it can be expressed in a simple analytic form.

Definition 18 (Tail Wasserstein Distance). *The p-th Wasserstein distance between F and G, conditioned on $[0, w]$, is defined to be*

$$\mathrm{WD}_p(F; G, w) \triangleq \left(\int_0^1 \left|F_w^{-1}(u) - G_w^{-1}(u)\right|^p \mathrm{d}u\right)^{\frac{1}{p}}.$$

For some of the aforementioned tail measures, we will also consider a normalization of the entropy, divergence or distance (as the case may be) with respect to w, the length of the tail. In Sections 5 and 6, we will show that as w tends to zero, the limits of these (possibly normalized) tail entropies and tail divergences can be expressed in terms of the local intrinsic dimensionalities of F and G. The notation for these variants, and our results for their limits in terms of ID_F^* and ID_G^*, are summarized in Table 1.

Table 1. Asymptotic equivalences between LID formulations and tail measures of entropy or divergence. In each case, the functions F and G are assumed to be smooth growth functions. In addition, for the Normalized Wasserstein Distance, F and G must be strictly monotonically increasing, thereby guaranteeing that the inverses of F_w and G_w exist near zero. In some cases, for the asymptotic limit to exist non-trivially (that is, to be both finite and non-zero), the tail entropy or tail divergence must be normalized by the multiplicative factor $\frac{1}{w}$, w. For the Tail Entropy and Tail Cross Entropy, no reweighting by powers of w can lead to a non-trivial asymptotic limit as w tends to zero.

Tail Measure	Formulation	Limit as $\mathbf{w \to 0^+}$
Entropy	$\mathrm{H}(F,w) = -\int_0^w F'_w(t) \ln F'_w(t)\,\mathrm{d}t$	Diverges (no reweighting possible)
Varentropy	$\mathrm{VarH}(F,w) = \int_0^w F'_w(t) \ln^2 F'_w(t)\,\mathrm{d}t - \left(\int_0^w F'_w(t) \ln F'_w(t)\,\mathrm{d}t\right)^2$	$\left(1 - \frac{1}{\mathrm{ID}_F^*}\right)^2$
q-Entropy	$\mathrm{H}_q(F,w) = \frac{1}{q-1}\int_0^w F'_w(t) - (F'_w(t))^q\,\mathrm{d}t$	$\frac{1}{q-1}$ if $q<1$, diverges if $q>1$
Normalized Cumulative Entropy	$\frac{1}{w}\mathrm{cH}(F,w) = -\frac{1}{w}\int_0^w F_w(t)\ln F_w(t)\,\mathrm{d}t$	$\frac{\mathrm{ID}_F^*}{(\mathrm{ID}_F^*+1)^2}$
Normalized Cumulative q-Entropy	$\frac{1}{w}\mathrm{cH}_q(F,w) = \frac{1}{w(q-1)}\int_0^w F_w(t) - (F_w(t))^q\,\mathrm{d}t$	$\frac{\mathrm{ID}_F^*}{(\mathrm{ID}_F^*+1)(q\,\mathrm{ID}_F^*+1)}$ if $q \neq 1$
Normalized Entropy Power	$\frac{1}{w}\mathrm{HP}(F,w) = \frac{1}{w}\exp(\mathrm{H}(F,w))$	$\frac{1}{\mathrm{ID}_F^*}\exp\left(1 - \frac{1}{\mathrm{ID}_F^*}\right)$
Normalized q-Entropy Power	$\frac{1}{w}\mathrm{HP}_q(F,w) = \frac{1}{w}\left[1 + (1-q)\,\mathrm{H}_q(F,w)\right]^{\frac{1}{1-q}}$	$\left(\frac{(\mathrm{ID}_F^*)^q}{q\,\mathrm{ID}_F^* - q + 1}\right)^{\frac{1}{1-q}}$ if $q \neq 1$ and $q\,\mathrm{ID}_F^* - q + 1 > 0$
Cross Entropy	$\mathrm{XH}(F;G,w) = -\int_0^w F'_w(t)\ln G'_w(t)\,\mathrm{d}t$	Diverges (no reweighting possible)
Normalized Cross Entropy Power	$\frac{1}{w}\mathrm{XHP}(F;G,w) = \frac{1}{w}\exp\left(-\int_0^w F'_w(t)\ln G'_w(t)\,\mathrm{d}t\right)$	$\frac{1}{\mathrm{ID}_G^*}\exp\left(\frac{\mathrm{ID}_G^* - 1}{\mathrm{ID}_F^*}\right)$
KL Divergence	$\mathrm{KL}(F;G,w) = \int_0^w F'_w(t)\ln\frac{F'_w(t)}{G'_w(t)}\,\mathrm{d}t$	$\rho - \ln\rho - 1;\ \rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$
JS Divergence	$\mathrm{JS}(F;G,w) = \frac{1}{2}\left(\mathrm{KL}\left(F;\frac{F+G}{2},w\right) + \mathrm{KL}\left(G;\frac{F+G}{2},w\right)\right)$	$\frac{1}{2}(\tau - \ln\tau - 1);\ \tau = \min\{\rho, \frac{1}{\rho}\};\ \rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$
Weighted L2 Distance	$w\,\mathrm{L2D}(F;G,w) = w\int_0^w (F'_w(t) - G'_w(t))^2\,\mathrm{d}t$	$\frac{\left(\mathrm{ID}_F^* - \mathrm{ID}_G^*\right)^2}{2(\mathrm{ID}_F^* + \mathrm{ID}_G^* - 1)}\left[1 + \frac{1}{(2\,\mathrm{ID}_F^* - 1)(2\,\mathrm{ID}_G^* - 1)}\right]$ $\mathrm{ID}_F^* > \frac{1}{2}\,;\mathrm{ID}_G^* > \frac{1}{2}$
Hellinger Distance	$\mathrm{HD}(F;G,w) = \sqrt{\frac{1}{2}\int_0^w \left(\sqrt{F'_w(t)} - \sqrt{G'_w(t)}\right)^2\,\mathrm{d}t}$	$\frac{\lvert 1-\sqrt{\rho}\rvert}{\sqrt{1+\rho}};\ \rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$
χ^2-Divergence	$\chi^2\mathrm{D}(F;G,w) = \int_0^w \frac{\left(F'_w(t) - G'_w(t)\right)^2}{G'_w(t)}\,\mathrm{d}t$	$\frac{(1-\rho)^2}{\rho(2-\rho)};\ \rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*};\ \rho < 2$
α-Divergence	$\alpha\mathrm{D}(F;G,w) = \frac{1}{\alpha(1-\alpha)}\int_0^w \alpha F'_w(t) + (1-\alpha)G'_w(t) - (F'_w(t))^\alpha (G'_w(t))^{1-\alpha}\,\mathrm{d}t$	$\frac{1}{\alpha(1-\alpha)}\left(1 - \frac{1}{\alpha\rho^{\alpha-1} + (1-\alpha)\rho^\alpha}\right)$ $\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}\ ;\ \alpha + \rho(1-\alpha) > 0$
Normalized Wasserstein Distance	$\frac{1}{w}\mathrm{WD}_p(F;G,w) = \frac{1}{w}\left(\int_0^1 \lvert F_w^{-1}(u) - G_w^{-1}(u)\rvert^p\,\mathrm{d}u\right)^{\frac{1}{p}}$	$p=2$: $\sqrt{\frac{1}{\frac{2}{\mathrm{ID}_F^*}+1} - \frac{2}{\frac{1}{\mathrm{ID}_F^*}+\frac{1}{\mathrm{ID}_G^*}+1} + \frac{1}{\frac{2}{\mathrm{ID}_G^*}+1}}$ p even: $\left(\sum_{j=0}^p \frac{(-1)^j\binom{p}{j}}{(p-j)\cdot(\mathrm{ID}_F^*)^{-1} + j\cdot(\mathrm{ID}_G^*)^{-1}+1}\right)^{\frac{1}{p}}$

5. Simplification of Tail Measures

Next, we present the main theoretical contributions of the paper: three technical lemmas that will later be used to establish relationships between local intrinsic dimensionality and a variety of tail measures based on entropy, divergences or distances. The results presented in this section all apply *asymptotically*, as the tail boundary tends toward zero.

Each of the three lemmas allow, under certain conditions, the simplification of limits of integrals involving smooth growth functions of the form F_w (as defined in Section 4), or

its associated first derivative F'_w or inverse function F_w^{-1}. The limit integral simplifications allow for the substitution of the function (or derivative or inverse) by expressions that involve one or more of the following: the LID value of the function, the variable of integration or the tail boundary w. Moreover, the lemmas require that the integrand be monotone with respect to small variations in the targeted function.

The first lemma allows terms of the form F_w (resembling the CDF of a tail-conditioned distribution) to be converted into a term that depends only on the variable of integration, the tail length w, and the local intrinsic dimension ID_F^*.

Lemma 1. *Let F be a smooth growth function over the interval $[0, r)$. Consider the function $\phi : \mathbb{R}_+^2 \to \mathbb{R}$ admitting a representation of the form*

$$\phi(t, w) \equiv \psi(t, w, z(t, w)),$$

where:

- $\psi : \mathbb{R}_+^3 \to \mathbb{R}$;
- $z(t, w) = F_w(t) = \frac{F(t)}{F(w)}$; *and*
- *for all fixed choices of t and w satisfying $0 < t \leq w < r$, $\psi(t, w, z)$ is monotone and continuously partially differentiable with respect to z over the interval $z \in (0, 1]$.*

Then

$$\begin{aligned} \lim_{w \to 0^+} \int_0^w \phi(t, w)\, \mathrm{d}t &\equiv \lim_{w \to 0^+} \int_0^w \psi(t, w, F_w(t))\, \mathrm{d}t \\ &= \lim_{w \to 0^+} \int_0^w \psi\left(t, w, \left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}\right) \mathrm{d}t \end{aligned}$$

whenever the latter limit exists or diverges to $+\infty$ or $-\infty$.

Proof. Since F is assumed to be a smooth growth function, the limit $\mathrm{ID}_F^* = \lim_{v \to 0^+} \mathrm{ID}_F(v)$ exists and is positive. We present an 'epsilon-delta' argument based on this limit. For any real value $\epsilon > 0$ satisfying $\epsilon < \min\{r, \mathrm{ID}_F^*\}$, there must exist a value $0 < \delta < \epsilon$ such that $v < \delta$ implies that $|\, \mathrm{ID}_F(v) - \mathrm{ID}_F^* \,| < \epsilon$. Therefore, when $0 < t \leq w < \delta$,

$$\left| \ln A_F(t, w) \right| = \left| \int_t^w \frac{\mathrm{ID}_F^* - \mathrm{ID}_F(v)}{v}\, \mathrm{d}v \right| < \epsilon \cdot \left| \int_t^w \frac{1}{v}\, \mathrm{d}v \right| = \epsilon \cdot \ln \frac{w}{t}\,.$$

Exponentiating, we obtain the bounds

$$\left(\frac{w}{t}\right)^{-\epsilon} < A_F(t, w) < \left(\frac{w}{t}\right)^{\epsilon}.$$

Applying this bound together with Theorem 2, the ratio $F_w(t) = \frac{F(t)}{F(w)}$ can be seen to satisfy

$$\left(\frac{t}{w}\right)^{\mathrm{ID}_F^* + \epsilon} < \frac{F(t)}{F(w)} = A_F(t, w) \cdot \left(\frac{t}{w}\right)^{\mathrm{ID}_F^*} < \left(\frac{t}{w}\right)^{\mathrm{ID}_F^* - \epsilon}. \tag{1}$$

Over the domain of interest $0 < t \leq w < \delta$, the assumption that $0 < \epsilon < \min\{r, \mathrm{ID}_F^*\}$ ensures that $0 < \frac{t}{w} \leq 1$, and that the upper and lower bounds of Inequality (1) lie in the interval $(0, 1]$. Since $\psi(t, w, z)$ has been assumed to be monotone with respect to $z \in (0, 1]$, the maximum and minimum attained by ψ over choices of z restricted to any (closed)

subinterval of $(0,1]$ must occur at opposite endpoints of the subinterval. With this in mind, for any choice of $\epsilon \in (0, \min\{r, \mathrm{ID}_F^*\})$, Inequality (1) implies that

$$\begin{aligned} B_{\min}(t,w,\epsilon) \;&\leq\; \psi(t,w,F_w(t)) \;\leq\; B_{\max}(t,w,\epsilon) \\ \text{and} \quad \int_0^w B_{\min}(t,w,\epsilon)\,\mathrm{d}t \;&\leq\; \int_0^w \psi(t,w,F_w(t))\,\mathrm{d}t \;\leq\; \int_0^w B_{\max}(t,w,\epsilon)\,\mathrm{d}t\,, \end{aligned}$$

where

$$\begin{aligned} B_{\min}(t,w,\epsilon) \;&\triangleq\; \min\left\{\psi\left(t,w,\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*-\epsilon}\right), \psi\left(t,w,\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*+\epsilon}\right)\right\}, \\ B_{\max}(t,w,\epsilon) \;&\triangleq\; \max\left\{\psi\left(t,w,\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*-\epsilon}\right), \psi\left(t,w,\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*+\epsilon}\right)\right\}. \end{aligned}$$

Since $\psi(t,w,z)$ and $\int_0^w \psi(t,w,z)\,\mathrm{d}t$ are also continuously partially differentiable with respect to z over $z \in (0,1]$,

$$\begin{aligned} \lim_{\epsilon\to 0^+} B_{\min}(t,w,\epsilon) \;&=\; \lim_{\epsilon\to 0^+} B_{\max}(t,w,\epsilon) \;=\; \psi\left(t,w,\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}\right) \quad \text{and} \\ \lim_{\substack{\epsilon\to 0^+ \\ w<\epsilon}} \int_0^w B_{\min}(t,w,\epsilon)\,\mathrm{d}t \;&=\; \lim_{\substack{\epsilon\to 0^+ \\ w<\epsilon}} \int_0^w B_{\max}(t,w,\epsilon)\,\mathrm{d}t \;=\; \lim_{w\to 0^+} \int_0^w \psi\left(t,w,\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}\right)\mathrm{d}t. \end{aligned}$$

It therefore follows from the squeeze theorem for integrals that

$$\lim_{w\to 0^+} \int_0^w \psi(t,w,F_w(t))\,\mathrm{d}t \;=\; \lim_{w\to 0^+} \int_0^w \psi\left(t,w,\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}\right)\mathrm{d}t\,,$$

whenever the right-hand limit exists or diverges. □

In a manner similar to that of the preceding lemma, the following result allows terms of the form F_w^{-1} (the inverse of F_w) to be converted into a term that depends only on the variable of integration, the tail length w and the local intrinsic dimension ID_F^*. Here, in order to ensure the existence of the inverse function, F (and by extension F_w and F_w^{-1}) must be strictly monotonically increasing over the tail.

Lemma 2. *Let F be a smooth growth function over the interval $[0,r)$. Let us also assume that, over the interval, the monotonicity of F is strict. Consider the function $\phi : \mathbb{R}_+^2 \to \mathbb{R}$ admitting a representation of the form*

$$\phi(u,w) \;\equiv\; \psi(u,w,z(u,w)),$$

where:

- *$\psi : \mathbb{R}_+^3 \to \mathbb{R}$;*
- *$z(u,w) = F_w^{-1}(u)$ for all $w \in (0,r)$, where $F_w(t) \triangleq F(t)/F(w)$ is restricted to values of t in $[0,w]$; and*
- *for all fixed choices of u and w satisfying $u \in [0,1]$ and $0 < w < r$, $\psi(u,w,z)$ is monotone and continuously partially differentiable with respect to z over the interval $z \in (0,r)$.*

Then

$$\begin{aligned} \lim_{w\to 0^+} \int_0^1 \phi(u,w)\,\mathrm{d}u \;&\equiv\; \lim_{w\to 0^+} \int_0^1 \psi\left(u,w,F_w^{-1}(u)\right)\mathrm{d}u \\ &=\; \lim_{w\to 0^+} \int_0^1 \psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*}}\right)\mathrm{d}u \end{aligned}$$

whenever the latter limit exists or diverges to $+\infty$ or $-\infty$.

Proof. First, we note that the strict monotonicity of F implies that for all $u \in [0,1]$ and $w \in (0,r)$, the function $F_w^{-1}(u)$ is uniquely defined when F_w is restricted to $[0,w]$.

As in the proof of Lemma 1, an 'epsilon-delta' argument based on the existence of the limit $\mathrm{ID}_F^* = \lim_{v\to 0^+} \mathrm{ID}_F(v)$ yields the following: for any real value $\epsilon > 0$ satisfying $\epsilon < \min\{r, \mathrm{ID}_F^*\}$, there exists a value $\delta \in (0,\epsilon)$ such that

$$\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*+\epsilon} < F_w(t) = \frac{F(t)}{F(w)} < \left(\frac{t}{w}\right)^{\mathrm{ID}_F^*-\epsilon}$$

holds for all $0 < t \le w < \delta$. Solving for t through exponentiation of the bounds, and then setting $t = F_w^{-1}(u)$, we obtain

$$\begin{aligned} w\cdot(F_w(t))^{\frac{1}{\mathrm{ID}_F^*-\epsilon}} &< t < w\cdot(F_w(t))^{\frac{1}{\mathrm{ID}_F^*+\epsilon}} \\ w\cdot\left(F_w(F_w^{-1}(u))\right)^{\frac{1}{\mathrm{ID}_F^*-\epsilon}} &< F_w^{-1}(u) < w\cdot\left(F_w(F_w^{-1}(u))\right)^{\frac{1}{\mathrm{ID}_F^*+\epsilon}} \\ wu^{\frac{1}{\mathrm{ID}_F^*-\epsilon}} &< F_w^{-1}(u) < wu^{\frac{1}{\mathrm{ID}_F^*+\epsilon}}. \end{aligned}$$

The remainder of the proof follows essentially the same path as that of Lemma 1. Over the domain of interest $0 < t \le w < \delta$, the assumption that $0 < \epsilon < \min\{r, \mathrm{ID}_F^*\}$ ensures that $0 < \frac{t}{w} \le 1$, and that u lies in the interval $(0,w]$. Since $\psi(u,w,z)$ has been assumed to be monotone with respect to $z \in (0,r)$, the maximum and minimum attained by ψ over choices of z restricted to any (closed) subinterval of $(0,r)$ must occur at opposite endpoints. Therefore, for any choice of $\epsilon \in (0, \min\{r, \mathrm{ID}_F^*\})$,

$$\begin{aligned} & C_{\min}(u,w,\epsilon) \le \psi\left(u,w,F_w^{-1}(u)\right) \le C_{\max}(u,w,\epsilon) \\ \text{and} \quad & \int_0^1 C_{\min}(u,w,\epsilon)\,\mathrm{d}u \le \int_0^1 \psi\left(u,w,F_w^{-1}(u)\right)\mathrm{d}u \le \int_0^1 C_{\max}(u,w,\epsilon)\,\mathrm{d}u, \end{aligned}$$

where

$$\begin{aligned} C_{\min}(u,w,\epsilon) &\triangleq \min\left\{\psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*-\epsilon}}\right), \psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*+\epsilon}}\right)\right\}, \\ C_{\max}(u,w,\epsilon) &\triangleq \max\left\{\psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*-\epsilon}}\right), \psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*+\epsilon}}\right)\right\}. \end{aligned}$$

Since $\psi(u,w,z)$ is also continuously partially differentiable with respect to z over $z \in (0,r)$,

$$\begin{aligned} \lim_{\epsilon\to 0^+} C_{\min}(u,w,\epsilon) &= \lim_{\epsilon\to 0^+} C_{\max}(u,w,\epsilon) = \psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*}}\right) \quad \text{and} \\ \lim_{\substack{\epsilon\to 0^+ \\ w<\epsilon}} \int_0^1 C_{\min}(u,w,\epsilon)\,\mathrm{d}u &= \lim_{\substack{\epsilon\to 0^+ \\ w<\epsilon}} \int_0^1 C_{\max}(u,w,\epsilon)\,\mathrm{d}u = \lim_{w\to 0^+} \int_0^1 \psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*}}\right)\mathrm{d}u. \end{aligned}$$

It therefore follows from the squeeze theorem for integrals that

$$\lim_{w\to 0^+} \int_0^1 \psi\left(u,w,F_w^{-1}(u)\right)\mathrm{d}u = \lim_{w\to 0^+} \int_0^1 \psi\left(u,w,wu^{\frac{1}{\mathrm{ID}_F^*}}\right)\mathrm{d}u,$$

whenever the right-hand limit exists or diverges. □

The third lemma facilitates the conversion of a term of the form F_w' to F_w, together with a factor that depends only on the variable of integration and ID_F^*. Since F is assumed to be a smooth growth function, F_w must be smooth as well, and therefore F_w satisfies

the conditions of Theorem 1 over $[0, w)$. Hence, F'_w can be substituted by an expression involving F_w:

$$F'_w(t) = \frac{\mathrm{ID}_{F_w}(t)}{t} \cdot F_w(t) = \frac{\mathrm{ID}_F(t)}{t} \cdot F_w(t)\,.$$

The substitution comes at the cost of introducing a non-constant factor $\mathrm{ID}_F(t)$. The following lemma shows that $\mathrm{ID}_F(t)$ can in turn be substituted by the constant ID_F^*, provided that certain monotonicity assumptions are satisfied.

Lemma 3. *Let F be a smooth growth function over the interval $[0, r)$. Consider the function $\phi : \mathbb{R}_+^2 \to \mathbb{R}$ admitting a representation of the form*

$$\phi(t, w) \;\equiv\; \psi(t, w, z(t, w)),$$

where:

- $\psi : \mathbb{R}_+^3 \to \mathbb{R}$*;*
- $z(t, w) = \mathrm{ID}_F(t)$*, and*
- *there exists a value $\gamma \in (0, \mathrm{ID}_F^*)$ such that for all fixed choices of t satisfying $0 < t \le w < r$, $\psi(t, w, z)$ is monotone with respect to z over the interval $z \in (\mathrm{ID}_F^* - \gamma, \mathrm{ID}_F^* + \gamma)$.*

Then

$$\lim_{w\to 0^+} \int_0^w \phi(t, w)\,\mathrm{d}t \quad\equiv\quad \lim_{w\to 0^+} \int_0^w \psi(t, w, \mathrm{ID}_F(t))\,\mathrm{d}t \;=\; \lim_{w\to 0^+} \int_0^w \psi(t, w, \mathrm{ID}_F^*)\,\mathrm{d}t$$

whenever the latter limit exists or diverges to $+\infty$ or $-\infty$.

Proof. Since F is assumed to be a smooth growth function, the limit $\mathrm{ID}_F^* = \lim_{v\to 0^+} \mathrm{ID}_F(v)$ exists and is positive. We present an 'epsilon-delta' argument based on this limit. For any real value $\epsilon > 0$ satisfying $\epsilon < \min\{r, \gamma\}$, there must exist a value $0 < \delta < \epsilon$ such that $v < \delta$ implies that $|\,\mathrm{ID}_F(v) - \mathrm{ID}_F^*\,| < \epsilon$.

Since $\psi(t, w, z)$ has been assumed to be monotone with respect to z over the interval $z \in (\mathrm{ID}_F^* - \gamma, \mathrm{ID}_F^* + \gamma)$, the restriction $v < \delta < \epsilon < \min\{r, \gamma\}$ ensures that $\psi(t, w, z)$ is monotone over the entire domain of interest $0 < t \le w < \delta$. Therefore, the maximum and minimum attained by ψ over choices of z restricted to any (closed) subinterval of $(\mathrm{ID}_F^* - \gamma, \mathrm{ID}_F^* + \gamma)$ must occur at opposite endpoints of the subinterval. As in the proof of Lemma 1,

$$\begin{aligned} D_{\min}(t, w, \epsilon) \;&\le\; \psi(t, w, \mathrm{ID}_F(t)) \;&\le\; D_{\max}(t, w, \epsilon) \\ \text{and}\quad \int_0^w D_{\min}(t, w, \epsilon)\,\mathrm{d}t \;&\le\; \int_0^w \psi(t, w, \mathrm{ID}_F(t))\,\mathrm{d}t \;&\le\; \int_0^w D_{\max}(t, w, \epsilon)\,\mathrm{d}t\,, \end{aligned}$$

where

$$\begin{aligned} D_{\min}(t, w, \epsilon) \;&\triangleq\; \min\{\psi(t, w, \mathrm{ID}_F^* - \epsilon), \psi(t, w, \mathrm{ID}_F^* + \epsilon)\}\,, \\ D_{\max}(t, w, \epsilon) \;&\triangleq\; \max\{\psi(t, w, \mathrm{ID}_F^* - \epsilon), \psi(t, w, \mathrm{ID}_F^* + \epsilon)\}\,. \end{aligned}$$

Since $\psi(t, w, z)$ is also continuously partially differentiable with respect to z over the range $(\mathrm{ID}_F^* - \gamma, \mathrm{ID}_F^* + \gamma)$,

$$\begin{aligned} \lim_{\epsilon\to 0^+} D_{\min}(t, w, \epsilon) \;&=\; \lim_{\epsilon\to 0^+} D_{\max}(t, w, \epsilon) \;&=\; \psi(t, w, \mathrm{ID}_F^*) \quad \text{and} \\ \lim_{\substack{\epsilon\to 0^+ \\ w<\epsilon}} \int_0^w D_{\min}(t, w, \epsilon)\,\mathrm{d}t \;&=\; \lim_{\substack{\epsilon\to 0^+ \\ w<\epsilon}} \int_0^w D_{\max}(t, w, \epsilon)\,\mathrm{d}t \;&=\; \lim_{w\to 0^+} \int_0^w \psi(t, w, \mathrm{ID}_F^*)\,\mathrm{d}t. \end{aligned}$$

It therefore follows from the squeeze theorem for integrals that

$$\lim_{w\to 0^+}\int_0^w \psi(t,w,\mathrm{ID}_F(t))\,\mathrm{d}t \;=\; \lim_{w\to 0^+}\int_0^w \psi(t,w,\mathrm{ID}_F^*)\,\mathrm{d}t\,,$$

whenever the right-hand limit exists or diverges. □

6. Derivation of the Limits of Tail Measures

In this section, we will see how the three substitution lemmas can be applied to the limits of tail measures of entropy, divergence or distance, so as to produce formulations that depend only on the local intrinsic dimensions of the functions involved. All three lemmas require that the integral function be monotone with respect to small variations in the term that is targeted for substitution. In the discussion, we choose two tail measures as running examples: the tail KL divergence and the second tail Wasserstein distance ($p = 2$).

6.1. Handling Derivatives of Smooth Growth Functions

In the case of the tail KL divergence, Theorem 1 allows us to substitute out the first derivatives F'_w and G'_w for the functions F_w and G_w:

$$\mathrm{KL}(F;G,w) \;=\; \int_0^w F'_w(t)\ln\frac{F'_w(t)}{G'_w(t)}\,\mathrm{d}t \;=\; \int_0^w \frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\ln\frac{\mathrm{ID}_F(t)\,F_w(t)}{\mathrm{ID}_G(t)\,G_w(t)}\,\mathrm{d}t\,.$$

6.2. Substitution of LID Functions by Constants

In the limit of the tail KL divergence, the functions $\mathrm{ID}_F(t)$ and $\mathrm{ID}_G(t)$ can be replaced by the constants ID_F^* and ID_G^*, respectively, through three successive applications of Lemma 3. To verify that the monotonicity condition of the Lemma is satisfied, we choose one of the terms and replace it by a new variable, z:

$$\lim_{w\to 0^+}\mathrm{KL}(F;G,w) \;=\; \lim_{w\to 0^+}\int_0^w \frac{z\,F_w(t)}{t}\ln\frac{\mathrm{ID}_F(t)\,F_w(t)}{\mathrm{ID}_G(t)\,G_w(t)}\,\mathrm{d}t\,.$$

For any fixed values of t and w, it is easy to see that the integrand is locally monotone in the vicinity of $z = \mathrm{ID}_F(t)$—here, if $\ln\frac{\mathrm{ID}_F(t)\,F_w(t)}{\mathrm{ID}_G(t)\,G_w(t)}$ is positive, a small increase in z (above the value $\mathrm{ID}_F(t)$) would result in an increase in the value of the integrand, and a small decrease would cause the integrand to decrease. If instead the logarithmic factor were negative, an increase in z would result in a decrease in the value of the integrand. Either way, the integrand would be monotone in the vicinity of $z = \mathrm{ID}_F(t)$ at each fixed value of t and w. Its monotonicity condition thus being satisfied, Lemma 3 allows the targeted instance of $\mathrm{ID}_F(t)$ to be substituted by ID_F^*:

$$\lim_{w\to 0^+}\mathrm{KL}(F;G,w) \;=\; \lim_{w\to 0^+}\int_0^w \frac{\mathrm{ID}_F^*\,F_w(t)}{t}\ln\frac{\mathrm{ID}_F(t)\,F_w(t)}{\mathrm{ID}_G(t)\,G_w(t)}\,\mathrm{d}t\,.$$

Similarly, it can be verified that the new integrand is monotone in each of the remaining two factors $\mathrm{ID}_F(t)$ and $\mathrm{ID}_G(t)$; consequently, they too can be substituted by ID_F^* and ID_G^*, one at a time, to yield

$$\lim_{w\to 0^+}\mathrm{KL}(F;G,w) \;=\; \lim_{w\to 0^+}\int_0^w \frac{\mathrm{ID}_F^*\,F_w(t)}{t}\ln\frac{\mathrm{ID}_F^*\,F_w(t)}{\mathrm{ID}_G^*\,G_w(t)}\,\mathrm{d}t\,.$$

6.3. Elimination of Tail-Conditioned Smooth Growth Functions

Now that the tail KL divergence has been reformulated in terms of the tail-conditioned smooth growth functions F_w and G_w, these two functions can be substituted out via three

successive applications of Lemma 1, so as to obtain the limit of an integral involving only the variable of integration t, and the constants w, ID_F and ID_G:

$$\lim_{w\to 0^+} \mathrm{KL}(F;G,w) \;=\; \lim_{w\to 0^+} \int_0^w \frac{\mathrm{ID}_F^*}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*} \ln\left[\frac{\mathrm{ID}_F^*}{\mathrm{ID}_G^*}\left(\frac{t}{w}\right)^{\mathrm{ID}_F^* - \mathrm{ID}_G^*}\right] \mathrm{d}t\,.$$

As in the previous step in which $\mathrm{ID}_F(t)$ and $\mathrm{ID}_G(t)$ were substituted out, the monotonicity conditions of Lemma 1 can easily be verified.

Now that the integral involves only constants and the variable t, it can be solved straightforwardly using the integration-by-parts technique, yielding

$$\lim_{w\to 0^+} \mathrm{KL}(F;G,w) \;=\; \lim_{w\to 0^+} \left(\frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - \ln\frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - 1\right) \;=\; \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - \ln\frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - 1\,.$$

6.4. Elimination of the Inverses of Tail-Conditioned Smooth Growth Functions

We now turn our attention to the limit of the tail Wasserstein distance for the case $p = 2$. Using Lemma 2, the inverse functions F_w^{-1} and G_w^{-1} can be substituted out, provided that the monotonicity requirements are satisfied. However, immediate application of the lemma to $F_w^{-1}(u)$ or $G_w^{-1}(u)$ does not necessarily work—to see this, consider substituting $F_w^{-1}(u)$ by the new variable z.

$$\mathrm{WD}_2(F;G,w) \;=\; \sqrt{\int_0^1 \left(F_w^{-1}(u) - G_w^{-1}(u)\right)^2 \mathrm{d}u} \;=\; \sqrt{\int_0^1 \left(z - G_w^{-1}(u)\right)^2 \mathrm{d}u}\,.$$

Clearly, the integrand is not necessarily monotone in z in the vicinity of those values of the integration variable u where $G_w^{-1}(u) = z$.

Instead, we expand the squared difference and apply Lemma 3 to each of the resulting four occurrences of F_w^{-1} and G_w^{-1}, one by one. By way of illustration, we consider substitution by z for the factor of $F_w^{-1}(u)$ in the cross term:

$$\begin{aligned}
\lim_{w\to 0^+} \left(\mathrm{WD}_2(F;G,w)\right)^2 &= \lim_{w\to 0^+} \int_0^1 \left(F_w^{-1}(u) - G_w^{-1}(u)\right)^2 \mathrm{d}u \\
&= \lim_{w\to 0^+} \int_0^1 \left(F_w^{-1}(u)\right)^2 - 2F_w^{-1}(u)G_w^{-1}(u) + \left(G_w^{-1}(u)\right)^2 \mathrm{d}u \\
&= \lim_{w\to 0^+} \int_0^1 \left(F_w^{-1}(u)\right)^2 - 2z\cdot G_w^{-1}(u) + \left(G_w^{-1}(u)\right)^2 \mathrm{d}u\,.
\end{aligned}$$

With respect to small variations in the variable z about the value $F_w^{-1}(u)$, noting that G_w^{-1} is always non-negative, the integrand is easily seen to be monotone in z when $G_w^{-1}(u)$ is non-zero: for any increase in z, the value of the integrand decreases, and for any decrease in z, the value of the integrand increases. Lemma 2 can therefore be applied, producing

$$\lim_{w\to 0^+} \left(\mathrm{WD}_2(F;G,w)\right)^2 \;=\; \lim_{w\to 0^+} \int_0^1 \left(F_w^{-1}(u)\right)^2 - 2wu^{\frac{1}{\mathrm{ID}_F^*}}\cdot G_w^{-1}(u) + \left(G_w^{-1}(u)\right)^2 \mathrm{d}u\,.$$

After three more applications of Lemma 2, followed by taking the square root of the integral, we obtain

$$\begin{aligned}
\lim_{w\to 0^+} \mathrm{WD}_2(F;G,w) &= \lim_{w\to 0^+} \left(\int_0^1 w^2u^{\frac{2}{\mathrm{ID}_F^*}} - 2w^2u^{\frac{1}{\mathrm{ID}_F^*}+\frac{1}{\mathrm{ID}_G^*}} + w^2u^{\frac{2}{\mathrm{ID}_G^*}}\,\mathrm{d}u\right)^{\frac{1}{2}} \\
&= \lim_{w\to 0^+} w\cdot\left(\frac{1}{\frac{2}{\mathrm{ID}_F^*}+1} - \frac{2}{\frac{1}{\mathrm{ID}_F^*}+\frac{1}{\mathrm{ID}_G^*}+1} + \frac{1}{\frac{2}{\mathrm{ID}_G^*}+1}\right)^{\frac{1}{2}} = 0\,.
\end{aligned}$$

6.5. Normalization

Even though the limit of the second tail Wasserstein distance is zero and therefore uninteresting, we observe that by normalizing it by the tail length w, we arrive at a more useful result:

$$\lim_{w\to 0^+} \frac{1}{w}\mathrm{WD}_2(F;G,w) \;=\; \left(\frac{1}{\frac{2}{\mathrm{ID}_F^*}+1} - \frac{2}{\frac{1}{\mathrm{ID}_F^*}+\frac{1}{\mathrm{ID}_G^*}+1} + \frac{1}{\frac{2}{\mathrm{ID}_G^*}+1}\right)^{\frac{1}{2}}.$$

In general, reweighting by a power of w may be required to expose a relationship between the tail limit of an entropy measure or divergence and an expression in terms of the local intrinsic dimensions of the functions involved. Since local intrinsic dimension is a unitless quantity, in order to establish a non-trivial formulation solely in terms of LID values, any tail measure whose values are not unitless will generally require some form of normalization.

6.6. Summary of Results

Table 1 provides a summary of results. All the results stated in Table 1 can be derived either using the techniques outlined earlier in this section, or through direct substitution of another result in the table. The derivations are outlined in Table 2 (tail entropy variants), Table 3 (tail divergence variants), Table 4 (tail distance variants) and Table 5 (tail Wasserstein distances). Most of these derivations are straightforward; however, for two of the tail measures, some clarifications are required.

Generally speaking, for the normalized tail Wasserstein distances with p non-integer or p odd (Table 5), Lemma 2 cannot be applied, due to the absolute value operation in the integrand. It may happen that the functions $F^{-1}(u)$ and $G^{-1}(u)$ may have crossing points for many (possibly even infinitely many) values of u between 0 and 1. At these values of u, $\left|F^{-1}(u) - G^{-1}(u)\right| = 0$, and neither $\left|z - G^{-1}(u)\right|$ nor $\left|F^{-1}(u) - y\right|$ would be monotone in the vicinity of $z = F^{-1}(u)$ or $y = G^{-1}(u)$, as the case may be.

For the tail JS divergence (Table 3), the derivation relies on the fact that the LID of the sum (or average) of two non-negative smooth growth functions is the smaller of the two individual LID values. This is an implication of the fact that $\lim_{t\to 0^+} \frac{V(t)}{W(t)} = 0$ whenever the smooth growth functions $V(t)$ and $W(t)$ have $0 < \mathrm{ID}_W^* < \mathrm{ID}_V^*$ (see [84] for more details). Accordingly, if $\mathrm{ID}_F^* \neq \mathrm{ID}_G^*$, then the function ($F$ or G) with smaller LID value must have the same LID value as the average function $M(t) = \frac{F(t)+G(t)}{2}$, and the other function (G or F) must have LID value equal to the maximum of the two. From these observations, the derivation can be seen to hold.

The result for the limit of the tail KL divergence has an interesting interpretation in light of the so-called Itakura–Saito (IS) divergence (or distance) [85]:

$$d_{\mathrm{IS}}(\mathbf{x}|\mathbf{y}) \;=\; \sum_{i=1}^{n}\left(\frac{x_i}{y_i} - \ln\frac{x_i}{y_i} - 1\right).$$

As the tail boundary w tends to 0, the tail KL divergence between smooth functions F and G tends to the (univariate) IS divergence between their associated LID values ID_G^* and ID_F^*:

$$\lim_{w\to 0^+} \mathrm{KL}(F;G,w) \;=\; \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - \ln\frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - 1 \;=\; d_{\mathrm{IS}}(\mathrm{ID}_G^* \,|\, \mathrm{ID}_F^*).$$

When F and G are interpreted as the CDFs of distance distributions, the shape parameters of the extreme-value-theoretic generalized pareto distributions (GPDs) that asymptotically characterize their lower tails are known to equal $-\frac{1}{\mathrm{ID}_F^*}$ and $-\frac{1}{\mathrm{ID}_G^*}$, respectively [40]. Since the ratio of these parameters is equal to (the reciprocal of) the ratio of LID val-

ues, the tail KL divergence between F and G can also be interpreted as tending to the IS divergence between GPD parameters.

Table 2. Derivations of asymptotic relationships between tail entropy variants and local intrinsic dimensionality. Each step shows the equivalences between the formulations when w is allowed to tend to zero. In the comments column, for each step of the derivation, the lemmas invoked are stated, as well as any additional assumptions made. If a normalization other weighting is needed to avoid divergence, or convergence to a constant (independent of F), the details are shown in a comment in the final step. In all cases, F is assumed to be a smooth growth function.

Tail Measure	Derivation Steps	Comments
Entropy	$\mathrm{H}(F,w) \to -\int_0^w F'_w(t)\ln F'_w(t)\,\mathrm{d}t$	
	$\to -\int_0^w \frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\ln\frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\,\mathrm{d}t$	using Theorem 1
	$\to -\int_0^w \frac{\mathrm{ID}^*_F\,F_w(t)}{t}\ln\frac{\mathrm{ID}^*_F\,F_w(t)}{t}\,\mathrm{d}t$	using Lemma 3
	$\to -\int_0^w \frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\ln\left[\frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\right]\mathrm{d}t$	using Lemma 1
	$\to 1-\frac{1}{\mathrm{ID}^*_F}-\ln\frac{\mathrm{ID}^*_F}{w}$	no reweighting
Varentropy	$\mathrm{VarH}(F,w) \to \int_0^w F'_w(t)\ln^2 F'_w(t)\,\mathrm{d}t - \left(\int_0^w F'_w(t)\ln F'_w(t)\,\mathrm{d}t\right)^2$	
	$\to \int_0^w \frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\ln^2\frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\,\mathrm{d}t - \left(\int_0^w \frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\ln\frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\,\mathrm{d}t\right)^2$	using Theorem 1
	$\to \int_0^w \frac{\mathrm{ID}^*_F\,F_w(t)}{t}\ln^2\frac{\mathrm{ID}^*_F\,F_w(t)}{t}\,\mathrm{d}t - \left(\int_0^w \frac{\mathrm{ID}^*_F\,F_w(t)}{t}\ln\frac{\mathrm{ID}^*_F\,F_w(t)}{t}\,\mathrm{d}t\right)^2$	using Lemma 3
	$\to \int_0^w \frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\ln^2\left[\frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\right]\mathrm{d}t - \left(\int_0^w \frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\ln\left[\frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\right]\mathrm{d}t\right)^2$	using Lemma 1
	$\to \left(1-\frac{1}{\mathrm{ID}^*_F}\right)^2$	
q-Entropy	$\mathrm{H}_q(F,w) \to \frac{1}{q-1}\int_0^w F'_w(t)-(F'_w(t))^q\,\mathrm{d}t$	$q>1$
	$\to \frac{1}{q-1}\int_0^w \frac{\mathrm{ID}_F(t)\,F_w(t)}{t}-\left(\frac{\mathrm{ID}_F(t)\,F_w(t)}{t}\right)^q\mathrm{d}t$	using Theorem 1
	$\to \frac{1}{q-1}\int_0^w \frac{\mathrm{ID}^*_F\,F_w(t)}{t}-\left(\frac{\mathrm{ID}^*_F\,F_w(t)}{t}\right)^q\mathrm{d}t$	using Lemma 3
	$\to \frac{1}{q-1}\int_0^w \frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}-\left(\frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\right)^q\mathrm{d}t$	using Lemma 1
	$\to \frac{1}{q-1}\left(1-\frac{1}{w^{q-1}}\cdot\frac{(\mathrm{ID}^*_F)^q}{q\,\mathrm{ID}^*_F-q+1}\right)$	
Cumulative Entropy	$\mathrm{cH}(F,w) \to -\int_0^w F_w(t)\ln F_w(t)\,\mathrm{d}t$	
	$\to -\int_0^w \left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\ln\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\mathrm{d}t$	using Lemma 1
	$\to w\frac{\mathrm{ID}^*_F}{(\mathrm{ID}^*_F+1)^2}$	weight by $\frac{1}{w}$
Cumulative q-Entropy	$\mathrm{cH}_q(F,w) \to \frac{1}{q-1}\int_0^w F_w(t)-(F_w(t))^q\,\mathrm{d}t$	$q\neq 1$
	$\to \frac{1}{q-1}\int_0^w \left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}-\left(\frac{t}{w}\right)^{q\,\mathrm{ID}^*_F}\mathrm{d}t$	using Lemma 1
	$\to w\frac{\mathrm{ID}^*_F}{(\mathrm{ID}^*_F+1)(q\,\mathrm{ID}^*_F+1)}$	weight by $\frac{1}{w}$
Entropy Power	$\mathrm{HP}(F,w) \to \exp(\mathrm{H}(F,w))$	
	$\to \exp\left(1-\frac{1}{\mathrm{ID}^*_F}-\ln\frac{\mathrm{ID}^*_F}{w}\right)$	by substitution
	$\to w\frac{1}{\mathrm{ID}^*_F}\exp\left(1-\frac{1}{\mathrm{ID}^*_F}\right)$	weight by $\frac{1}{w}$
q-Entropy Power	$\mathrm{HP}_q(F,w) \to \left[1+(1-q)\,\mathrm{H}_q(F,w)\right]^{\frac{1}{1-q}}$	$q\neq 1$
	$\to \left(1+(1-q)\cdot\frac{1}{q-1}\left[1-\frac{1}{w^{q-1}}\cdot\frac{(\mathrm{ID}^*_F)^q}{q\,\mathrm{ID}^*_F-q+1}\right]\right)^{\frac{1}{1-q}}$	by substitution
	$\to w\left(\frac{(\mathrm{ID}^*_F)^q}{q\,\mathrm{ID}^*_F-q+1}\right)^{\frac{1}{1-q}}$	weight by $\frac{1}{w}$

Table 3. Derivations of asymptotic relationships between tail divergences and local intrinsic dimensionality. Each step shows the equivalences between the formulations when w is allowed to tend to zero. In the comments column, for each step of the derivation, the lemmas invoked are stated, as well as any additional assumptions made. If a normalization or weighting is needed, the details are shown in a comment in the final step. In all cases, F and G are assumed to be smooth growth functions.

Tail Measure	Derivation Steps	Comments
Cross Entropy	$\mathrm{XH}(F;G,w) \to -\int_0^w F'_w(t) \ln G'_w(t)\,\mathrm{d}t$	
	$\to -\int_0^w \frac{\mathrm{ID}_F(t)\,F_w(t)}{t} \ln \frac{\mathrm{ID}_G(t)\,G_w(t)}{t}\,\mathrm{d}t$	using Theorem 1
	$\to -\int_0^w \frac{\mathrm{ID}^*_F\,F_w(t)}{t} \ln \frac{\mathrm{ID}^*_G\,G_w(t)}{t}\,\mathrm{d}t$	using Lemma 3
	$\to -\int_0^w \frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F} \ln\left[\frac{\mathrm{ID}^*_G}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_G}\right]\mathrm{d}t$	using Lemma 1
	$\to \frac{\mathrm{ID}^*_G - 1}{\mathrm{ID}^*_F} - \ln \frac{\mathrm{ID}^*_G}{w}$	no reweighting
Cross Entropy Power	$\mathrm{XHP}(F;G,w) \to \exp(\mathrm{XH}(F;G,w))$	
	$\to \exp\left(\frac{\mathrm{ID}^*_G - 1}{\mathrm{ID}^*_F} - \ln \frac{\mathrm{ID}^*_G}{w}\right)$	by substitution
	$\to w^{\frac{1}{\mathrm{ID}^*_G}} \exp\left(\frac{\mathrm{ID}^*_G - 1}{\mathrm{ID}^*_F}\right)$	weight by $\frac{1}{w}$
KL Divergence	$\mathrm{KL}(F;G,w) \to \int_0^w F'_w(t) \ln \frac{F'_w(t)}{G'_w(t)}\,\mathrm{d}t$	
	$\to \int_0^w \frac{\mathrm{ID}_F(t)\,F_w(t)}{t} \ln \frac{\mathrm{ID}_F(t)\,F_w(t)}{\mathrm{ID}_G(t)\,G_w(t)}\,\mathrm{d}t$	using Theorem 1
	$\to \int_0^w \frac{\mathrm{ID}^*_F\,F_w(t)}{t} \ln \frac{\mathrm{ID}^*_F\,F_w(t)}{\mathrm{ID}^*_G\,G_w(t)}\,\mathrm{d}t$	using Lemma 3
	$\to \int_0^w \frac{\mathrm{ID}^*_F}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F} \ln\left[\frac{\mathrm{ID}^*_F}{\mathrm{ID}^*_G}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F - \mathrm{ID}^*_G}\right]\mathrm{d}t$	using Lemma 1
	$\to \rho - \ln\rho - 1$	$\rho = \frac{\mathrm{ID}^*_G}{\mathrm{ID}^*_F}$
JS Divergence	$\mathrm{JS}(F;G,w) \to \frac{1}{2}(\mathrm{KL}(F;M,w) + \mathrm{KL}(G;M,w))$	$M(t) = \frac{1}{2}(F(t) + G(t))$
	$\to \frac{1}{2}\left(\frac{\mathrm{ID}^*_M}{\mathrm{ID}^*_F} - \ln\frac{\mathrm{ID}^*_M}{\mathrm{ID}^*_F} - 1 + \frac{\mathrm{ID}^*_M}{\mathrm{ID}^*_G} - \ln\frac{\mathrm{ID}^*_M}{\mathrm{ID}^*_G} - 1\right)$	$\mathrm{ID}^*_M = \min\{\mathrm{ID}^*_F, \mathrm{ID}^*_G\}$
	$\to \frac{1}{2}\left(\frac{\mathrm{ID}^*_M}{B} + \frac{\mathrm{ID}^*_M}{\mathrm{ID}^*_M} - \ln\frac{\mathrm{ID}^*_M}{B} - \ln\frac{\mathrm{ID}^*_M}{\mathrm{ID}^*_M} - 2\right)$	let $B = \max\{\mathrm{ID}^*_F, \mathrm{ID}^*_G\}$
	$\to \frac{1}{2}(\tau - \ln\tau - 1)$	$\tau = \min\left\{\frac{\mathrm{ID}^*_G}{\mathrm{ID}^*_F}, \frac{\mathrm{ID}^*_F}{\mathrm{ID}^*_G}\right\}$

The IS divergence is popular as an objective for matrix factorization of audio spectra [86], for assessing the loss of using entry $y_{i,j}$ to approximate a true entry $x_{i,j}$; more precisely, to approximate a matrix $\mathbf{V}$ by factorization $\mathbf{WH}$, the loss is $\sum_i \sum_j d_{\mathrm{IS}}([\mathbf{V}]_{ij}|[\mathbf{WH}]_{ij})$. The IS divergence is a convenient choice for this scenario due to its scale-free property ($d_{\mathrm{IS}}(\mathbf{x}|\mathbf{y}) = d_{\mathrm{IS}}(\alpha\mathbf{x}|\alpha\mathbf{y})$ for any $\alpha \neq 0$), thus giving the same relative weight to both small and large values of x_i and y_i, since they only appear as the ratio $\frac{x_i}{y_i}$. This is important for scenarios such as audio spectra, where the magnitudes of x_i and y_i can vary greatly.

The Itakura–Saito divergence falls into the family of so-called Bregman divergences (or distances) [87], which have a geometric interpretation as the difference between the value of a convex generator function at $\mathbf{x}$ on the one hand, and the value at $\mathbf{x}$ of a hyperplane function that is tangent to the generator curve at $\mathbf{y}$. Bregman divergences are a highly expressive family of distances with a wide range of applications [88]. For the IS divergence, the convex generator function is the negative logarithm $-\sum_{i=1}^n \ln x_i$. Interestingly, the KL

divergence is also a Bregman divergence, with its convex generator being the negative entropy function $\sum_{i=1}^{n} x_i \ln x_i$ [89].

Table 4. Derivations of asymptotic relationships between tail distances and local intrinsic dimensionality. Each step shows the equivalences between the formulations when w is allowed to tend to zero. In the comments column, for each step of the derivation, the lemmas invoked are stated, as well as any additional assumptions made. For each tail distance, the first step of the derivations shows an expansion by which the monotonicity of each factor can be verified. If a normalization or weighting is needed, the details are shown in a comment in the final step. In all cases, F and G are assumed to be smooth growth functions.

Tail Measure	Derivation Steps	Comments
L2 Distance	$\mathrm{L2D}(F;G,w) \to \int_0^w (F'_w(t) - G'_w(t))^2 \,\mathrm{d}t$	
	$\to \int_0^w \left(\frac{\mathrm{ID}_F(t)\, F_w(t)}{t} - \frac{\mathrm{ID}_G(t)\, G_w(t)}{t} \right)^2 \mathrm{d}t$	using Theorem 1
	$\to \int_0^w \left(\frac{\mathrm{ID}_F^*\, F_w(t)}{t} \right)^2 - 2\frac{\mathrm{ID}_F^*\, F_w(t)}{t} \cdot \frac{\mathrm{ID}_G^*\, G_w(t)}{t} + \left(\frac{\mathrm{ID}_G^*\, G_w(t)}{t} \right)^2 \mathrm{d}t$	using Lemma 3
	$\to \int_0^w \frac{(\mathrm{ID}_F^*)^2}{t^2} \left(\frac{t}{w}\right)^{2\,\mathrm{ID}_F^*} - \frac{2\,\mathrm{ID}_F^*\,\mathrm{ID}_G^*}{t^2} \left(\frac{t}{w}\right)^{\mathrm{ID}_F^* + \mathrm{ID}_G^*} + \frac{(\mathrm{ID}_G^*)^2}{t^2} \left(\frac{t}{w}\right)^{2\,\mathrm{ID}_G^*} \mathrm{d}t$	using Lemma 1
	$\to \frac{1}{w} \cdot \frac{(\mathrm{ID}_F^* - \mathrm{ID}_G^*)^2}{2(\mathrm{ID}_F^* + \mathrm{ID}_G^* - 1)} \left[1 + \frac{1}{(2\,\mathrm{ID}_F^* - 1)(2\,\mathrm{ID}_G^* - 1)} \right]$	weight by w
Hellinger Distance	$\mathrm{HD}(F;G,w) \to \sqrt{\frac{1}{2} \int_0^w \left(\sqrt{F'_w(t)} - \sqrt{G'_w(t)} \right)^2 \mathrm{d}t}$	
	$\to \sqrt{\frac{1}{2} \int_0^w \left(\sqrt{\frac{\mathrm{ID}_F(t)\, F_w(t)}{t}} - \sqrt{\frac{\mathrm{ID}_G(t)\, G_w(t)}{t}} \right)^2 \mathrm{d}t}$	using Theorem 1
	$\to \sqrt{\frac{1}{2} \int_0^w \frac{\mathrm{ID}_F^*\, F_w(t)}{t} - 2\frac{\sqrt{\mathrm{ID}_F^*\, F_w(t) \cdot \mathrm{ID}_G^*\, G_w(t)}}{t} + \frac{\mathrm{ID}_G^*\, G_w(t)}{t} \,\mathrm{d}t}$	using Lemma 3
	$\to \sqrt{\int_0^w \frac{1}{2t} \left(\mathrm{ID}_F^* \left(\frac{t}{w}\right)^{\mathrm{ID}_F^*} - 2\sqrt{\mathrm{ID}_F^*\,\mathrm{ID}_G^*} \left(\frac{t}{w}\right)^{(\mathrm{ID}_F^* + \mathrm{ID}_G^*)/2} + \mathrm{ID}_G^* \left(\frac{t}{w}\right)^{\mathrm{ID}_G^*} \right) \mathrm{d}t}$	using Lemma 1
	$\to \frac{\lvert 1 - \sqrt{\rho} \rvert}{\sqrt{1+\rho}}$	$\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$
χ^2-Divergence	$\chi^2\mathrm{D}(F;G,w) \to \int_0^w \frac{\left(F'_w(t) - G'_w(t)\right)^2}{G'_w(t)} \,\mathrm{d}t$	
	$\to \int_0^w \left(\frac{\mathrm{ID}_F(t)\, F_w(t)}{t} - \frac{\mathrm{ID}_G(t)\, G_w(t)}{t} \right)^2 \frac{t}{\mathrm{ID}_G(t)\, G_w(t)} \,\mathrm{d}t$	using Theorem 1
	$\to \int_0^w \left[\left(\frac{\mathrm{ID}_F^*\, F_w(t)}{t} \right)^2 - 2\frac{\mathrm{ID}_F^*\, F_w(t)}{t} \cdot \frac{\mathrm{ID}_G^*\, G_w(t)}{t} + \left(\frac{\mathrm{ID}_G^*\, G_w(t)}{t} \right)^2 \right] \frac{t}{\mathrm{ID}_G^*\, G_w(t)} \,\mathrm{d}t$	using Lemma 3
	$\to \int_0^w \left[\frac{(\mathrm{ID}_F^*)^2}{t^2} \left(\frac{t}{w}\right)^{2\,\mathrm{ID}_F^*} - \frac{2\,\mathrm{ID}_F^*\,\mathrm{ID}_G^*}{t^2} \left(\frac{t}{w}\right)^{\mathrm{ID}_F^* + \mathrm{ID}_G^*} + \frac{(\mathrm{ID}_G^*)^2}{t^2} \left(\frac{t}{w}\right)^{2\,\mathrm{ID}_G^*} \right] \frac{t}{\mathrm{ID}_G^*} \left(\frac{w}{t}\right)^{\mathrm{ID}_G^*} \mathrm{d}t$	using Lemma 1
	$\to \frac{(1-\rho)^2}{\rho(2-\rho)}$	$\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$
α-Divergence	$\alpha\mathrm{D}(F;G,w) \to \frac{1}{\alpha(1-\alpha)} \int_0^w \alpha F'_w(t) + (1-\alpha) G'_w(t) - (F'_w(t))^\alpha (G'_w(t))^{1-\alpha} \,\mathrm{d}t$	
	$\to \frac{1}{\alpha(1-\alpha)} \int_0^w \frac{\alpha\, \mathrm{ID}_F(t)\, F_w(t)}{t} + \frac{(1-\alpha)\, \mathrm{ID}_G(t)\, G_w(t)}{t} - \left(\frac{\mathrm{ID}_F(t)\, F_w(t)}{t} \right)^\alpha \left(\frac{\mathrm{ID}_G(t)\, G_w(t)}{t} \right)^{1-\alpha} \mathrm{d}t$	using Theorem 1
	$\to \frac{1}{\alpha(1-\alpha)} \int_0^w \frac{\alpha\, \mathrm{ID}_F^*\, F_w(t)}{t} + \frac{(1-\alpha)\, \mathrm{ID}_G^*\, G_w(t)}{t} - \left(\frac{\mathrm{ID}_F^*\, F_w(t)}{t} \right)^\alpha \left(\frac{\mathrm{ID}_G^*\, G_w(t)}{t} \right)^{1-\alpha} \mathrm{d}t$	using Lemma 3
	$\to \frac{1}{\alpha(1-\alpha)} \int_0^w \frac{\alpha\, \mathrm{ID}_F^*}{t} \left(\frac{t}{w}\right)^{\mathrm{ID}_F^*} + \frac{(1-\alpha)\, \mathrm{ID}_G^*}{t} \left(\frac{t}{w}\right)^{\mathrm{ID}_G^*} - \frac{(\mathrm{ID}_F^*)^\alpha (\mathrm{ID}_G^*)^{1-\alpha}}{t} \left(\frac{t}{w}\right)^{\alpha\,\mathrm{ID}_F^* + (1-\alpha)\,\mathrm{ID}_G^*} \mathrm{d}t$	using Lemma 1
	$\to \frac{1}{\alpha(1-\alpha)} \left(1 - \frac{(\mathrm{ID}_F^*)^\alpha (\mathrm{ID}_G^*)^{1-\alpha}}{\alpha\,\mathrm{ID}_F^* + (1-\alpha)\,\mathrm{ID}_G^*} \right)$	
	$\to \frac{1}{\alpha(1-\alpha)} \left(1 - \frac{1}{\alpha\rho^{\alpha-1} + (1-\alpha)\rho^\alpha} \right)$	$\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$

Table 5. Derivations of asymptotic relationships between tail Wasserstein distances and local intrinsic dimensionality. Each step shows the equivalences between the formulations when w is allowed to tend to zero. In the comments column, for each step of the derivation, the lemmas invoked are stated, as well as any additional assumptions made. Normalization details are shown in a comment in the final step. In all cases, F and G are assumed to be invertible smooth growth functions.

Tail Measure	Derivation Steps	Comments
Wasserstein Distance	$\mathrm{WD}_2(F;G,w) \to \sqrt{\int_0^1 \left(F_w^{-1}(u) - G_w^{-1}(u)\right)^2 \mathrm{d}u}$	
	$\to \sqrt{\int_0^1 \left(F_w^{-1}(u)\right)^2 - 2F_w^{-1}(u)\cdot G_w^{-1}(u) + \left(G_w^{-1}(u)\right)^2 \mathrm{d}u}$	
$p = 2$	$\to \sqrt{\int_0^1 w^2 u^{\frac{2}{\mathrm{ID}_F^*}} - 2w^2 u^{\frac{1}{\mathrm{ID}_F^*}+\frac{1}{\mathrm{ID}_G^*}} + w^2 u^{\frac{2}{\mathrm{ID}_G^*}} \mathrm{d}u}$	using Lemma 2
	$\to w\sqrt{\frac{1}{\frac{2}{\mathrm{ID}_F^*}+1} - \frac{2}{\frac{1}{\mathrm{ID}_F^*}+\frac{1}{\mathrm{ID}_G^*}+1} + \frac{1}{\frac{2}{\mathrm{ID}_G^*}+1}}$	weight by $\frac{1}{w}$
Wasserstein Distance	$\mathrm{WD}_p(F;G,w) \to \left(\int_0^1 \left(F_w^{-1}(u) - G_w^{-1}(u)\right)^p \mathrm{d}u\right)^{\frac{1}{p}}$	
$p \in \mathbb{N}$, p even	$\to \left(\int_0^1 \sum_{j=0}^{p} (-1)^j \binom{p}{j} \left(F_w^{-1}(u)\right)^{p-j} \left(G_w^{-1}(u)\right)^j \mathrm{d}u\right)^{\frac{1}{p}}$	
	$\to \left(\int_0^1 \sum_{j=0}^{p} (-1)^j \binom{p}{j} \left(w u^{\frac{1}{\mathrm{ID}_F^*}}\right)^{p-j} \left(w u^{\frac{1}{\mathrm{ID}_G^*}}\right)^j \mathrm{d}u\right)^{\frac{1}{p}}$	using Lemma 2
	$\to w\left(\sum_{j=0}^{p} \frac{(-1)^j \binom{p}{j}}{(p-j)\cdot(\mathrm{ID}_F^*)^{-1} + j\cdot(\mathrm{ID}_G^*)^{-1}+1}\right)^{\frac{1}{p}}$	weight by $\frac{1}{w}$

7. Extension to Multivariate Distributions

Thus far, our results have focused on a univariate scenario, wherein entropy and divergence variants were shown to be asymptotically equivalent to formulations involving the local intrinsic dimensionalities of smooth distributions of a single random variable. As discussed in Section 3, these results can be applied to distance-based analysis, through characterizations involving the LIDs of local (univariate) distance distributions induced by the overall (global) multivariate distribution. These characterizations are indirect, in that they do not explicitly involve (nor do they require) any knowledge of the underlying global distribution and its parameters. However, characterizations in terms of induced distance distributions may not be entirely satisfying when the nature of the global multivariate distribution is either known or assumed. In this section, we will assume that our domain $\mathcal{S}$ is the n-dimensional space $\mathbb{R}^n$ equipped with the Euclidean distance, $d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$. Within $\mathcal{S}$, we will also assume that we are given a data distribution $\mathcal{D}$ with probability density function $p : \mathbb{R}^n \to \mathbb{R}_+ \cup 0$.

7.1. Multivariate Tail Distributions with Local Spherical Symmetry

Within the Euclidean domain, the challenge is to analyze distributions in terms of the probability measure captured within volumes associated with a distributional tail. However, unlike in univariate distributions, there is no universally accepted notion of 'distributional tail' for multivariate distributions. For our purposes, given a distance $r > 0$, we define the tail of $\mathcal{D}$ of length r to be the region enclosed by the ball of radius r centered at the origin; that is, $\mathcal{B}(r) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq r\}$. The boundary of the tail is the $(n-1)$-dimensional surface area of $\mathcal{B}(r)$, which we denote by $\mathcal{B}'(r) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = r\}$.

To enable tractable analysis, we will assume that the PDF can be expressed in terms of a locally spherically symmetrical function. One example of where local spherical symmetry can be expected to hold is for a locally isotropic context. This is a common assumption for

physical systems, including metals, glasses, fluids and polymers, for which the distribution locally surrounding a particle in the system does not have a directional preference.

Formally, we say that a density function f is locally spherically symmetrical within radius w if for all $\|\mathbf{x}\| \leq w$, we have $f(\mathbf{x}) = f_*(r)$ for some univariate function f_* where $r = \|\mathbf{x}\|$. For f to be locally spherically symmetrical, it suffices that $f(\mathbf{x})$ be equal to $f(\mathbf{y})$ whenever $0 \leq \|\mathbf{x}\| = \|\mathbf{y}\| \leq r$. The assumption also implies the existence of a function f_* for which $f(\mathbf{x}) = f_*(r)$, and therefore that f must be constant over all points of the sphere $\mathcal{B}'(r)$.

The probability measure captured by $\mathcal{B}(r)$, which we denote by $F(r)$, is obtained through the integration of f over this ball:

$$F(r) \triangleq \int_{\mathcal{B}(r)} f \, \mathrm{d}\mathcal{B}(r) \, .$$

It is not difficult to see that the univariate function F is simply the CDF of the distribution of distances to the origin induced by the global distribution $\mathcal{D}$. If F is differentiable over the tail interval $(0, r]$, then the integral of F' over this interval exists, and equals F:

$$\int_{\mathcal{B}(r)} f \, \mathrm{d}\mathcal{B}(r) \quad = F(r) = \quad \int_0^r F'(t) \, \mathrm{d}t \, . \tag{2}$$

The derivative $F'(r)$ can therefore be interpreted as the PDF of the radial distance distribution as measured from the origin.

For spherically symmetric distributions in Euclidean spaces, the multivariate density and radial density is related through a factor that depends on the surface area of spherical volumes. The formulae for the volume of an n-sphere and its $(n-1)$-dimensional surface area are given by

$$\begin{aligned} V_n(r) &\triangleq \frac{\pi^{n/2}}{\Gamma((n/2)+1)} r^n \quad \text{and} \\ S_{n-1}(r) &\triangleq \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} \, , \end{aligned}$$

respectively. Γ is the common gamma function $\Gamma(n) = (n-1)!$ if n is a positive integer and $\Gamma(n + \frac{1}{2}) = (n - \frac{1}{2})(n - \frac{3}{2}) \ldots \frac{1}{2}\sqrt{\pi}$ if n is a non negative integer. Furthermore, the volume and surface area have a simple relationship that allows for easy conversion between the two:

$$r \cdot S_{n-1}(r) \; = \; n \cdot V_n(r) \; . \tag{3}$$

Lemma 4 ([90]). *Let* $\mathbf{X}$ *be an n-dimensional random vector that is spherically symmetric with a radial distribution* $\mathcal{R}$*. Then* $\mathbf{X}$ *has a density* $f(\mathbf{x})$ *if and only if* $\mathcal{R}$ *has a density s and*

$$s(r) = f(\mathbf{x}) \cdot S_{n-1}(r) \, .$$

If F is a smooth growth function that is locally spherically symmetric over $[0, r]$, Equation (2) and Lemma 4 together give us the following relationship between the radial density F' and the multivariate density f:

$$f(\mathbf{x}) \; = \; \frac{F'(\|\mathbf{x}\|)}{S_{n-1}(\|\mathbf{x}\|)}$$

whenever $\|\mathbf{x}\| \leq r$. Conditioning the distribution to the ball $\mathcal{B}(r)$, the tail distribution PDF becomes

$$f_r(\mathbf{x}) \triangleq \frac{f(\mathbf{x})}{\int_{\mathcal{B}(w)} f \, \mathrm{d}\mathcal{B}(w)} = \frac{F'(\|\mathbf{x}\|)}{S_{n-1}(\|\mathbf{x}\|) \cdot F(r)} = \frac{F'_r(\|\mathbf{x}\|)}{S_{n-1}(\|\mathbf{x}\|)} \, .$$

7.2. Multivariate Tail Entropy Variants

The aforementioned relationships between multivariate and radial densities can be immediately used to compute the various tail entropies for the locally spherically symmetric multivariate case. Useful background on evaluating radial integrals can be found in Baker [91]. For example, the multivariate Tail Entropy is

$$\begin{aligned}\mathrm{H}(f,w) &\triangleq -\int_{\mathcal{B}(w)} f_w \ln f_w \,\mathrm{d}\mathcal{B}(w)\\ &= -\int_0^w \left(\frac{F'_w(t)}{S_{n-1}(t)} \ln \frac{F'_w(t)}{S_{n-1}(t)}\right) \cdot S_{n-1}(t)\,\mathrm{d}t\\ &= -\int_0^w F'_w(t) \ln \frac{F'_w(t)}{S_{n-1}(t)}\,\mathrm{d}t\,.\end{aligned}$$

Although the multivariate formulation of Tail Entropy $\mathrm{H}(f,w)$ resembles that of the univariate formulation $\mathrm{H}(F,w)$, the two are not identical. Nevertheless, the multivariate formulation can still be simplified using the technical lemmas introduced in Section 5. In much the same way as for the univariate Tail Entropy Power, we can use Theorem 1 together with Lemmas 1 and 3 to determine the limit of $\mathrm{H}(f,w)$ as w tends to 0. Replacing $F'_w(t)$ by $\frac{1}{t}\,\mathrm{ID}_F(t)F_w(t)$, then $\mathrm{ID}_F(t)$ by ID_F^*, and finally $F_w(t)$ by $\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}$, we obtain

$$\begin{aligned}\lim_{w\to 0}\mathrm{H}(f,w) &= \lim_{w\to 0} -\int_0^w F'_w(t)\ln\frac{F'_w(t)}{t^{n-1}S_{n-1}(1)}\,\mathrm{d}t\\ &= \lim_{w\to 0} -\int_0^w \frac{\mathrm{ID}_F^*}{t}\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}\ln\left[\frac{\mathrm{ID}_F^*}{t^n S_{n-1}(1)}\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}\right]\mathrm{d}t\\ &= \lim_{w\to 0} -\int_0^w \frac{\mathrm{ID}_F^*}{w^{\mathrm{ID}_F^*}}t^{\mathrm{ID}_F^*-1}\ln\left[\frac{\mathrm{ID}_F^*}{w^{\mathrm{ID}_F^*}S_{n-1}(1)}t^{\mathrm{ID}_F^*-n}\right]\mathrm{d}t\\ &= \lim_{w\to 0} -\int_0^w \frac{\mathrm{ID}_F^*}{w^{\mathrm{ID}_F^*}}t^{\mathrm{ID}_F^*-1}\left[\ln\frac{\mathrm{ID}_F^*}{w^{\mathrm{ID}_F^*}S_{n-1}(1)} + (\mathrm{ID}_F^*-n)\ln t\right]\mathrm{d}t\,.\end{aligned}$$

Solving the integral, and then using Equation (3) to convert the surface area factor S_{n-1} to an expression involving the volume V_n, we eventually arrive at

$$\begin{aligned}\lim_{w\to 0}\mathrm{H}(f,w) &= \lim_{w\to 0}\left(1-\frac{n}{\mathrm{ID}_F^*}-\ln\frac{\mathrm{ID}_F^*}{w^n\,S_{n-1}(1)}\right)\\ &= \lim_{w\to 0}\left(1-\frac{n}{\mathrm{ID}_F^*}-\ln\frac{\mathrm{ID}_F^*}{w\,S_{n-1}(w)}\right)\\ &= \lim_{w\to 0}\left(1-\frac{n}{\mathrm{ID}_F^*}-\ln\frac{\mathrm{ID}_F^*}{n}+\ln V_n(w)\right),\end{aligned}$$

which diverges even when the Tail Entropy is reweighted by $V_n(w)$ (or indeed, by any other polynomial in w). However, the Tail Entropy Power, when normalized by $V_n(w)$, does converge to a strictly positive value:

$$\begin{aligned}\lim_{w\to 0}\frac{1}{V_n(w)}\mathrm{HP}(f,w) &\triangleq \lim_{w\to 0}\frac{1}{V_n(w)}\exp(\mathrm{H}(f,w))\\ &= \lim_{w\to 0}\frac{1}{V_n(w)}\exp\left(1-\frac{n}{\mathrm{ID}_F^*}-\ln\frac{\mathrm{ID}_F^*}{n}+\ln V_n(w)\right)\\ &= \frac{1}{\varphi}\exp\left(1-\frac{1}{\varphi}\right),\quad \text{where } \varphi=\frac{\mathrm{ID}_F^*}{n}\,.\end{aligned}$$

As one might expect in the n-dimensional Euclidean setting, the (normalized asymptotic) multivariate Tail Entropy Power is maximized whenever ID_F^*, the local intrinsic dimensionality of the associated radial CDF F, is equal to n.

7.3. Multivariate Cumulative Tail Entropy

In the multivariate setting, cumulative entropy is defined in terms of the distributional tail, according to the notion laid out in Section 7.1. In place of the usual probability density $f(\mathbf{x})$, the entropy function is applied to the probability measure associated with the ball centered at the origin with radius $\|\mathbf{x}\|$; that is, with

$$\Pr[X \leq \|\mathbf{x}\|] \triangleq \int_{\mathcal{B}(\|\mathbf{x}\|)} f \, \mathrm{d}\mathcal{B}(\|\mathbf{x}\|) = F(\|\mathbf{x}\|) \,.$$

Note that since F takes the same value at $\mathbf{x}$ and $\mathbf{y}$ whenever $\|\mathbf{x}\| = \|\mathbf{y}\|$, the quantity $F(\|\mathbf{x}\|)$ is locally spherically symmetric even when the underlying density function f is not.

We can adapt the multivariate formulation of cumulative residual entropy that was originally proposed by Rao [56]. The multivariate Cumulative Tail Entropy, conditioned to a distributional tail of radius w, is expressed as a multivariate integral involving $F_w(\|\mathbf{x}\|)$, or as a radial integral involving F_w, as follows:

$$\begin{aligned} \mathrm{cH}(f, w) &\triangleq -\int_{\mathbf{x} \in \mathcal{B}(w)} F_w(\|\mathbf{x}\|) \, \ln F_w(\|\mathbf{x}\|) \, \mathrm{d}\mathcal{B}(w) \\ &= -\int_0^w (F_w(t) \, \ln F_w(t)) \cdot S_{n-1}(t) \, \mathrm{d}t \,. \end{aligned}$$

As in the treatment of the univariate tail entropies, we can use Lemma 1 to determine the limit of $\mathrm{cH}(f, w)$ as w tends to 0. Replacing $F_w(t)$ by $\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*}$,

$$\begin{aligned} \lim_{w \to 0} \mathrm{cH}(f, w) &= \lim_{w \to 0} -\int_0^w t^{n-1} S_{n-1}(1) \, F_w(t) \, \ln F_w(t) \, \mathrm{d}t \\ &= \lim_{w \to 0} -\int_0^w t^{n-1} S_{n-1}(1) \left(\frac{t}{w}\right)^{\mathrm{ID}_F^*} \ln\left(\frac{t}{w}\right)^{\mathrm{ID}_F^*} \mathrm{d}t \\ &= \lim_{w \to 0} -\int_0^w \frac{S_{n-1}(1) \, \mathrm{ID}_F^*}{w^{\mathrm{ID}_F^*}} \cdot t^{\mathrm{ID}_F^* + n - 1} \, (\ln t - \ln w) \, \mathrm{d}t \,. \end{aligned}$$

Solving the integral, and then converting the surface area factor S_{n-1} to a volume factor V_n using Equation (3), we obtain

$$\begin{aligned} \lim_{w \to 0} \mathrm{cH}(f, w) &= \lim_{w \to 0} w \, S_{n-1}(w) \cdot \frac{\mathrm{ID}_F^*}{(\mathrm{ID}_F^* + n)^2} \\ &= \lim_{w \to 0} V_n(w) \cdot \frac{\varphi}{(\varphi + 1)^2} \,, \quad \text{where } \varphi = \frac{\mathrm{ID}_F^*}{n} \,. \end{aligned}$$

Although the multivariate Cumulative Tail Entropy vanishes as the tail boundary w tends to zero, when normalized by the tail volume $V_n(w)$ it converges to a strictly positive value:

$$\lim_{w \to 0} \frac{1}{V_n(w)} \mathrm{cH}(f, w) = \frac{\varphi}{(\varphi + 1)^2} \,.$$

Again, as with the Normalized Tail Entropy Power, the (asymptotic) multivariate Tail Cumulative Entropy is maximized whenever $\varphi = 1$. That is, when $\mathrm{ID}_F^* = n$.

7.4. Multivariate Tail Divergences

Several of the tail divergence measures, when considered in the multivariate setting under the assumptions of locally spherical symmetry, turn out to be identical to those of the

radial (univariate) setting. As an example, consider the multivariate Tail KL Divergence, defined as

$$\mathrm{KL}(f;g,w) \triangleq \int_{\mathcal{B}(w)} f_w \ln \frac{f_w}{g_w} \,\mathrm{d}\mathcal{B}(w)\,.$$

Applying Lemma 4 and integrating radially over the tail, we see that

$$\begin{aligned}\mathrm{KL}(f;g,w) &= \int_0^w \left(\frac{F'_w(t)}{S_{n-1}(t)} \ln \frac{F'_w(t)/S_{n-1}(t)}{G'_w(t)/S_{n-1}(t)}\right) \cdot S_{n-1}(t)\,\mathrm{d}t \\ &= \int_0^w F'_w(t) \ln \frac{F'_w(t)}{G'_w(t)}\,\mathrm{d}t \\ &= \mathrm{KL}(F;G,w)\,,\end{aligned}$$

the Tail KL Divergence of F and G, which (as stated in Table 1) has the limit $\frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - \ln \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} - 1$ as the tail length w tends to zero.

Similarly, it can easily be seen that the multivariate versions of the JS Divergence, the Hellinger Distance, the χ^2-Divergence and the α-Divergence all have radial integral formulations identical to their corresponding univariate versions.

7.5. Observations

The general strategy for deriving these results is essentially the same as for the multivariate Tail Entropy: first use Lemma 4 to convert the multidimensional integral to an integral in one dimension, then use the technical lemmas of Section 5 to simplify the univariate integral as before.

Our results for the locally spherically symmetric multivariate case are shown in Table 6; however, since their derivations greatly resemble those of the analogous univariate cases, we omit the details. Some remarks:

1. A result for the Wasserstein Distance is not included, since its formulation does not generalize straightforwardly to higher dimensions, unlike the other divergence measures.
2. The normalizations and weightings used depend only on the tail volume $V_n(w)$ and (for the Tsallis entropy variants) the parameter q. This generalizes our earlier univariate results where normalization was performed with regard to the tail length w.
3. All the multivariate tail variants considered Table 6 are elegant generalizations of their corresponding univariate formulations, and all explicitly depend on the ratios between the LIDs and the dimension of the space n ($\varphi = \frac{\mathrm{ID}_F^*}{n}$ and $\gamma = \frac{\mathrm{ID}_G^*}{n}$), or on the ratio of two LID values ($\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} = \frac{\gamma}{\varphi}$). Among these, the Normalized Entropy Power and the Normalized Cumulative Entropy are maximized when $\mathrm{ID}_F^* = n$, which can occur when the tail distribution is uniform. The Varentropy is minimized when $\mathrm{ID}_F^* = n$, which can occur when the variance of the log-likelihood for a uniform distribution is equal to zero.
4. As mentioned in Related Work, a number of previous studies in deep learning have found that the local intrinsic dimension in learned representations is lower than the dimension of the full space [32–35] (i.e., $\mathrm{ID}_F^* < n$) and that the learning process progressively reduces local intrinsic dimension. Consider a concrete example where $n = 100$ and $\mathrm{ID}_F^* = 12$ and the learning process is reducing ID_F^* at a point from 12 to 11. The consequent effect on entropy can be interpreted from two different perspectives, either as an increase in tail distance entropy or a decrease in tail location entropy:
   - Considering univariate normalized entropy power or normalized cumulative entropy (Table 1), reduction of ID_F^* corresponds to an increase in entropy. Here, the entropy is measuring the uncertainty of the univariate random variable

modeling distances to nearest neighbors. Thus, reduction of ID_F^* corresponds to an increase in "distance entropy".
- Considering multivariate normalized entropy power or multivariate normalized cumulative entropy (Table 6), reduction of ID_F^* corresponds to an decrease in entropy. Here, the entropy is measuring the uncertainty of the multivariate random variable modeling locations of nearest neighbors, assuming local spherical symmetry. So reduction of ID_F^* corresponds to a decrease in "location entropy".

We will see a visualization of these scenarios in Section 7.6.

5. All four of the multivariate tail divergences listed in Table 6, as well as the Hellinger Distance, have radial integral formulations that are identical to their univariate counterparts. All the divergences and distances (including the Weighted L2 Distance) are minimized when $\mathrm{ID}_F^* = \mathrm{ID}_G^*$.
6. By setting $n = 1$, we can recover the univariate results from Table 1. However, note that the range of integration used in Table 6 is a hypersphere of radius w, where for $n = 1$ it is the interval $[-w, w]$. In contrast, the integral formulations listed in Table 1 were taken over the interval $[0, w]$. For some results, this means a minor (constant factor of 2) difference between Table 1 and the result from Table 6 when $n = 1$.

Table 6. Asymptotic equivalences between LID formulations and tail measures of entropy or divergence for locally spherically symmetric distributions in the n-dimensional Euclidean setting. In each case, the density functions are assumed to be f and g, and the CDFs F and G of their induced distance distributions are assumed to be smooth growth functions. In the results, $V_n(r)$ and $S_{n-1}(r)$ denote the volume and surface area of the n-dimensional ball with radius r (respectively). In some cases, for the asymptotic limit to exist non-trivially (that is, to be both finite and non-zero), the tail entropy or tail divergence must be normalized by some multiplicative factor dependent on the tail volume $V_n(w)$.

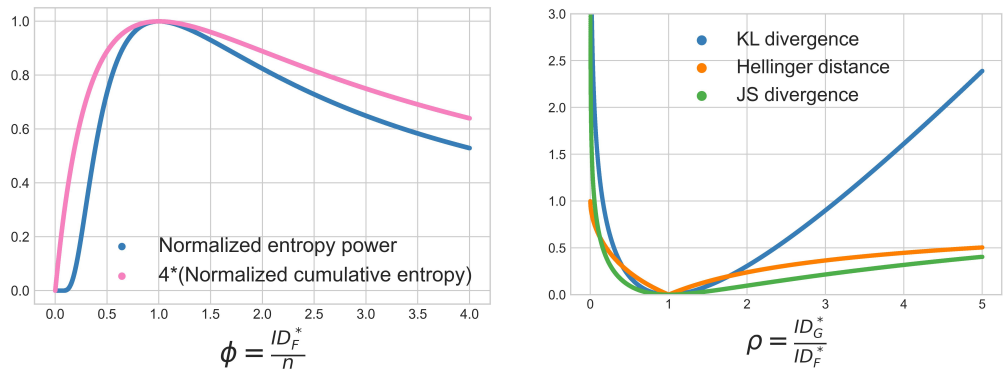
Tail Measure	**Formulation**	**Limit as w $\to 0^+$**
Entropy	$\mathrm{H}(f,w) = -\int_{\mathcal{B}(w)} f_w \ln f_w \,\mathrm{d}\mathcal{B}(w) = -\int_0^w F'_w(t) \ln \frac{F'_w(t)}{S_{n-1}(t)} \,\mathrm{d}t$	Diverges (no reweighting possible)
Varentropy	$\mathrm{VarH}(f,w) = \int_{\mathcal{B}(w)} f_w \ln^2 f_w \,\mathrm{d}\mathcal{B}(w) - \left(\int_{\mathcal{B}(w)} f_w \ln f_w \,\mathrm{d}\mathcal{B}(w)\right)^2$ $= \int_0^w F'_w(t) \ln^2 \frac{F'_w(t)}{S_{n-1}(t)} \,\mathrm{d}t - \left(\int_0^w F'_w(t) \ln \frac{F'_w(t)}{S_{n-1}(t)} \,\mathrm{d}t\right)^2$	$\left(1 - \frac{1}{\varphi}\right)^2$ $\varphi = \frac{\mathrm{ID}_F^*}{n}$
q-Entropy	$\mathrm{H}_q(f,w) = \frac{1}{q-1} \int_{\mathcal{B}(w)} f_w - f_w^q \,\mathrm{d}\mathcal{B}(w)$ $= \frac{1}{q-1} \int_0^w F'_w(t) - \frac{(F'_w(t))^q}{(S_{n-1}(t))^{q-1}} \,\mathrm{d}t$	$\frac{1}{q-1}$ if $q < 1$ diverges if $q > 1$
Normalized Cumulative Entropy	$\frac{1}{V_n(w)}\mathrm{cH}(f,w) = -\frac{1}{V_n(w)} \int_{\mathbf{x}\in\mathcal{B}(w)} F_w(\|\mathbf{x}\|) \ln F_w(\|\mathbf{x}\|) \,\mathrm{d}\mathcal{B}(w)$ $= -\frac{1}{V_n(w)} \int_0^w (F_w(t) \ln F_w(t)) \cdot S_{n-1}(t) \,\mathrm{d}t$	$\frac{\varphi}{(\varphi+1)^2}$ $\varphi = \frac{\mathrm{ID}_F^*}{n}$
Normalized Cumulative q-Entropy	$\frac{1}{V_n(w)}\mathrm{cH}_q(f,w) = -\frac{1}{V_n(w)} \cdot \frac{1}{q-1} \int_{\mathbf{x}\in\mathcal{B}(w)} F_w(\|\mathbf{x}\|) - (F_w(\|\mathbf{x}\|))^q \,\mathrm{d}\mathcal{B}(w)$ $= \frac{1}{V_n(w)} \cdot \frac{1}{q-1} \int_0^w \left(F_w(t) - (F_w(t))^q\right) \cdot S_{n-1}(t) \,\mathrm{d}t$	$\frac{\varphi}{(q\varphi+1)(\varphi+1)}$ if $q \neq 1$ $\varphi = \frac{\mathrm{ID}_F^*}{n}$
Normalized Entropy Power	$\frac{1}{V_n(w)}\mathrm{HP}(f,w) = \frac{1}{V_n(w)} \exp(\mathrm{H}(f,w))$	$\frac{1}{\varphi} \exp\left(1 - \frac{1}{\varphi}\right)$; $\varphi = \frac{\mathrm{ID}_F^*}{n}$
Normalized q-Entropy Power	$\frac{1}{V_n(w)}\mathrm{HP}_q(f,w) = \frac{1}{V_n(w)} \left[1 + (1-q)\,\mathrm{H}_q(f,w)\right]^{\frac{1}{1-q}}$	$\left(\frac{\varphi^q}{q\varphi - q + 1}\right)^{\frac{1}{1-q}}$; $\varphi = \frac{\mathrm{ID}_F^*}{n}$ if $q \neq 1$ and $q\varphi - q + 1 > 0$
Cross Entropy	$\mathrm{XH}(f;g,w) = -\int_{\mathcal{B}(w)} f_w \ln g_w \,\mathrm{d}\mathcal{B}(w) = -\int_0^w F'_w(t) \ln \frac{G'_w(t)}{S_{n-1}(t)} \,\mathrm{d}t$	Diverges (no reweighting possible)
Normalized Cross Entropy Power	$\frac{1}{V_n(w)}\mathrm{XHP}(f;g,w) = \frac{1}{V_n(w)} \exp(\mathrm{XH}(f;g,w))$	$\frac{1}{\gamma} \exp\left(\frac{\gamma-1}{\varphi}\right)$; $\varphi = \frac{\mathrm{ID}_F^*}{n}$; $\gamma = \frac{\mathrm{ID}_G^*}{n}$

Table 6. *Cont.*

Tail Measure	Formulation	Limit as $w \to 0^+$
Weighted L2 Distance	$V_n(w) \cdot \mathrm{L2D}(f;g,w) = V_n(w) \int_{\mathcal{B}(w)} (f_w - g_w)^2 \, \mathrm{d}\mathcal{B}(w)$ $= V_n(w) \int_0^w \frac{1}{S_{n-1}(t)} (F'_w(t) - G'_w(t))^2 \, \mathrm{d}t$	$\frac{(\varphi-\gamma)^2}{2(\varphi+\gamma-1)} \left[1 + \frac{1}{(2\varphi-1)(2\gamma-1)}\right]$ $\varphi = \frac{\mathrm{ID}_F^*}{n}$; $\gamma = \frac{\mathrm{ID}_G^*}{n}$ $\mathrm{ID}_F^* > \frac{1}{2}$; $\mathrm{ID}_G^* > \frac{1}{2}$
Hellinger Distance	$\mathrm{HD}(f;g,w) = \sqrt{\frac{1}{2} \int_{\mathcal{B}(w)} \left(\sqrt{f_w} - \sqrt{g_w}\right)^2 \mathrm{d}\mathcal{B}(w)}$ $= \sqrt{\frac{1}{2} \int_0^w \left(\sqrt{F'_w(t)} - \sqrt{G'_w(t)}\right)^2 \mathrm{d}t}$	$\frac{\lvert 1-\sqrt{\rho} \rvert}{\sqrt{1+\rho}}$ $\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$
χ^2-Divergence	$\chi^2\mathrm{D}(f;g,w) = \int_{\mathcal{B}(w)} \frac{(f_w - g_w)^2}{g_w} \, \mathrm{d}\mathcal{B}(w)$ $= \int_0^w \frac{\left(F'_w(t) - G'_w(t)\right)^2}{G'_w(t)} \, \mathrm{d}t$	$\frac{(1-\rho)^2}{\rho(2-\rho)}$ $\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$; $\rho < 2$
α-Divergence	$\alpha\mathrm{D}(f;g,w) = \frac{1}{\alpha(1-\alpha)} \int_{\mathcal{B}(w)} \alpha f_w + (1-\alpha) g_w - f_w^\alpha g_w^{1-\alpha} \, \mathrm{d}\mathcal{B}(w)$ $= \frac{1}{\alpha(1-\alpha)} \int_0^w \alpha F'_w(t) + (1-\alpha) G'_w(t)$ $- (F'_w(t))^\alpha (G'_w(t))^{1-\alpha} \, \mathrm{d}t$	$\frac{1}{\alpha(1-\alpha)} \left(1 - \frac{1}{\alpha\rho^{\alpha-1} + (1-\alpha)\rho^\alpha}\right)$ $\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$ Require $\alpha + \rho(1-\alpha) > 0$
KL Divergence	$\mathrm{KL}(f;g,w) = \int_{\mathcal{B}(w)} f_w \ln \frac{f_w}{g_w} \, \mathrm{d}\mathcal{B}(w) = \int_0^w F'_w(t) \ln \frac{F'_w(t)}{G'_w(t)} \, \mathrm{d}t$	$\rho - \ln\rho - 1$; $\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$
JS Divergence	$\mathrm{JS}(f;g,w) = \frac{1}{2}\left(\mathrm{KL}\left(f; \frac{f+g}{2}, w\right) + \mathrm{KL}\left(g; \frac{f+g}{2}, w\right)\right)$	$\frac{\tau - \ln\tau - 1}{2}$; $\tau = \min\{\rho, \frac{1}{\rho}\}$; $\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$

7.6. Visualization of Behavior

Our results in Table 6 relate local intrinsic dimensionality to entropies and divergences. If analyzing an n dimensional global distribution such as the standard normal distribution or uniform distribution, then the dimension of every sub-manifold (i.e., the local intrinsic dimensionality ID_F^*) will be n. However, our interest is in situations where the local intrinsic dimensionality differs from the representation dimension n. To provide further intuition on this aspect, two plots are shown in Figure 1.



(**a**) (**b**)

Figure 1. Visualization of selected measures from Table 6 (**a**) Entropy behavior as the ratio $\frac{\mathrm{ID}_F^*}{n}$ varies; (**b**) Divergence/distance behavior as the ratio $\frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$ varies.

Figure 1a compares the behavior of the normalized entropy power and the normalized cumulative entropy (multiplied by a constant factor of 4) in n-dimensional space, as the ratio $\phi = \frac{\mathrm{ID}_F^*}{n}$ is varied. We see that these measures have similar trends and they are maximized when $\mathrm{ID}_F^* = n$. We also see that when $1 \ll \mathrm{ID}_F^* < n$, these entropic measures will decrease if ID_F^* is decreased (for a fixed n). On the other hand, if $n = 1$ and $1 \ll \mathrm{ID}_F^*$, then these entropic measures will increase if ID_F^* is decreased, where $n = 1$ corresponds

to the scenario where we are modeling the uncertainty of a distance distribution. This illustrates remark number 4 from Section 7.5 above.

Figure 1b compares the behavior of different tail divergences as the ratio $\rho = \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*}$ varies. The divergences shown are the KL divergence, the Jensen–Shannon divergence and the Hellinger distance. These measures have similar trends as ρ varies and are minimized and equal to zero when $\mathrm{ID}_F^* = \mathrm{ID}_G^*$. Also, the Hellinger distance is bounded above by 1.

8. Conclusions

In this theoretical investigation, we have established asymptotic relationships between tail entropy variants, tail divergences and the theory of local intrinsic dimensionality. Our results are derived under the assumption that the distribution(s) under consideration are being analyzed in a highly local context, within the distribution tail(s), an asymptotically small neighborhood whose radius approaches zero. These results show that tail entropies and tail divergences depend in a fundamental way on local intrinsic dimensionality and help form a theoretical foundation for cross-fertilization between intrinsic dimensionality research and entropy research. As future work, we plan to investigate the potential of these new characterizations in a range of application settings. For example, for use as a basis in machine learning to characterize and improve representations and representation learning, as well as use in understanding behavior of physical systems such as fluids and helping characterize their critical transitions in time and space.

Our results from both univariate and multivariate cases, show that the tail entropies and divergences considered in this paper depend only on (i) the embedding (representation) dimension in which the distribution is situated, and (ii) the local intrinsic dimension(s) of the distribution(s). Furthermore, in many cases there is dependence involving the ratio between the intrinsic dimension and the embedding dimension.

Consider the context of distance based analysis, when a distribution models distances from a central query location to its nearest neighbors, and the distances are induced by global data. In this situation, our characterization of entropy might be termed as 'personalized', in that entropy expresses the uncertainty (or complexity) from the perspective of the query, in regard to the distances to samples within an asymptotically small neighborhood. Phrased another way, these local entropies are 'observer-dependent', since they are tied to the choice of query (the observer). This can be contrasted with the more common notion of entropy, where one analyzes a global distribution, and there is no requirement of a query point or its local neighborhood.

As alluded to in the introduction, divergences between tail distributions could be used for comparison of real and synthetic distributions, as is commonly required for generative adversarial networks (GANs). Given a particular query location we may either: (i) compute the divergence between the univariate tail distance distributions of synthetic and real examples, as measured from a query point; or (ii) compute the divergence between the multivariate tail distance distributions of synthetic and real examples, again as measured from the query, under an assumption of local isotropy. Our results show that under the assumption of local spherical symmetry, the use of divergences (such as KL) between tail distance distributions is asymptotically equivalent to the standard multivariate formulations with the same divergences, when restricted to the neighborhoods around locations of interest. For future work it will be interesting to consider whether it is possible to further extend our multivariate results to elliptically symmetric distributions or skew-elliptical distributions, such as those studied by Contreras-Reyes [65].

Lastly, our results in Tables 1 and 6 show theoretical relationships for entropies and divergences, but in practice one must estimate the measures using samples of data. A natural approach here is to first estimate local intrinsic dimensional values such as ID_F^* and ID_G^* using any desired estimator (such as the maximum likelihood estimator [39–41]), and then plug in the estimated LID value into the desired tail entropy or tail divergence formula. For example, an estimator of the (univariate) Normalized Cumulative Entropy

could be obtained by computing $\frac{\widehat{\mathrm{ID}_F^*}}{(\widehat{\mathrm{ID}_F^*}+1)^2}$, where $\widehat{\mathrm{ID}_F^*}$ is the estimated LID of the distance distribution F.

Author Contributions: Conceptualization, J.B., M.E.H. and X.M.; methodology, J.B., M.E.H. and X.M.; formal analysis, J.B., M.E.H. and X.M.; writing—original draft preparation, J.B., M.E.H. and X.M.; writing—review and editing, J.B., M.E.H. and X.M. All authors have read and agreed to the published version of the manuscript.

Funding: James Bailey acknowledges the support of ARC Discovery Grant DP170102472. Michael E. Houle acknowledges the financial support of JSPS Kakenhi Kiban (B) Research Grant 18H03296.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Basseville, M. Divergence measures for statistical data processing—An annotated bibliography. *Signal Process.* **2013**, *93*, 621–633. [CrossRef]
2. Houle, M.E. Local Intrinsic Dimensionality I: An Extreme-Value-Theoretic Foundation for Similarity Applications. In Proceedings of the International Conference on Similarity Search and Applications, Munich, Germany, 4–6 October 2017; pp. 64–79.
3. Bailey, J.; Houle, M.E.; Ma, X. Relationships Between Local Intrinsic Dimensionality and Tail Entropy. In Proceedings of the Similarity Search and Applications—Proc. of the 14th International Conference, SISAP 2021, Dortmund, Germany, 29 September–1 October 2021.
4. Heller, R.; Heller, Y. Multivariate tests of association based on univariate tests. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*; Lee, D.D., Sugiyama, M., von Luxburg, U., Guyon, I., Garnett, R., Eds.; Curran Associates Inc.: Red Hook, NY, USA, 2016; pp. 208–216.
5. Maa, J.; Pearl, D.; Bartoszynski, R. Reducing multidimensional two-sample data to one-dimensional interpoint comparisons. *Ann. Stat.* **1996**, *24*, 1069–1074. [CrossRef]
6. Li, A.; Qi, J.; Zhang, R.; Ma, X.; Ramamohanarao, K. Generative image inpainting with submanifold alignment. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, Hong Kong, 10–16 August 2019; pp. 811–817.
7. Camastra, F.; Staiano, A. Intrinsic dimension estimation: Advances and open problems. *Inf. Sci.* **2016**, *328*, 26–41. [CrossRef]
8. Campadelli, P.; Casiraghi, E.; Ceruti, C.; Rozza, A. Intrinsic Dimension Estimation: Relevant Techniques and a Benchmark Framework. *Math. Probl. Eng.* **2015**, *2015*, 759567. [CrossRef]
9. Verveer, P.J.; Duin, R.P.W. An evaluation of intrinsic dimensionality estimators. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *17*, 81–86. [CrossRef]
10. Bruske, J.; Sommer, G. Intrinsic dimensionality estimation with optimally topology preserving maps. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 572–575. [CrossRef]
11. Pettis, K.W.; Bailey, T.A.; Jain, A.K.; Dubes, R.C. An intrinsic dimensionality estimator from near-neighbor information. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 25–37. [CrossRef]
12. Navarro, G.; Paredes, R.; Reyes, N.; Bustos, C. An empirical evaluation of intrinsic dimension estimators. *Inf. Syst.* **2017**, *64*, 206–218. [CrossRef]
13. Jolliffe, I.T. *Principal Component Analysis*; Springer: Berlin/Heidelberg, Germany, 2002.
14. Costa, J.A.; Hero III, A.O. Entropic Graphs for Manifold Learning. In Proceedings of the 37th Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 1, pp. 316–320.
15. Hein, M.; Audibert, J.Y. Intrinsic dimensionality estimation of submanifolds in R^d. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 289–296.
16. Rozza, A.; Lombardi, G.; Rosa, M.; Casiraghi, E.; Campadelli, P. IDEA: Intrinsic Dimension Estimation Algorithm. In Proceedings of the International Conference on Image Analysis and Processing, Ravenna, Italy, 14–16 September 2011; pp. 433–442.
17. Rozza, A.; Lombardi, G.; Ceruti, C.; Casiraghi, E.; Campadelli, P. Novel High Intrinsic Dimensionality Estimators. *Mach. Learn.* **2012**, *89*, 37–65. [CrossRef]
18. Ceruti, C.; Bassis, S.; Rozza, A.; Lombardi, G.; Casiraghi, E.; Campadelli, P. DANCo: An intrinsic dimensionality estimator exploiting angle and norm concentration. *Pattern Recognit.* **2014**, *47*, 2569–2581. [CrossRef]
19. Facco, E.; d'Errico, M.; Rodriguez, A.; Laio, A. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Sci. Rep.* **2017**, *7*, 12140. [CrossRef]
20. Zhou, S.; Tordesillas, A.; Pouragha, M.; Bailey, J.; Bondell, H. On local intrinsic dimensionality of deformation in complex materials. *Nat. Sci. Rep.* **2021**, *11*, 10216. [CrossRef]

21. Tordesillas, A.; Zhou, S.; Bailey, J.; Bondell, H. A representation learning framework for detection and characterization of dead versus strain localization zones from pre- to post- failure. *Granul. Matter* **2022**, *24*, 75. [CrossRef]
22. Faranda, D.; Messori, G.; Yiou, P. Dynamical proxies of North Atlantic predictability and extremes. *Sci. Rep.* **2017**, *7*, 41278. [CrossRef]
23. Messori, G.; Harnik, N.; Madonna, E.; Lachmy, O.; Faranda, D. A dynamical systems characterization of atmospheric jet regimes. *Earth Syst. Dynam.* **2021**, *12*, 233–251. [CrossRef]
24. Kambhatla, N.; Leen, T.K. Dimension Reduction by Local Principal Component Analysis. *Neural Comput.* **1997**, *9*, 1493–1516. [CrossRef]
25. Houle, M.E.; Ma, X.; Nett, M.; Oria, V. Dimensional Testing for Multi-Step Similarity Search. In Proceedings of the IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 299–308.
26. Campadelli, P.; Casiraghi, E.; Ceruti, C.; Lombardi, G.; Rozza, A. Local Intrinsic Dimensionality Based Features for Clustering. In Proceedings of the International Conference on Image Analysis and Processing, Naples, Italy, 9–13 September 2013; pp. 41–50.
27. Houle, M.E.; Schubert, E.; Zimek, A. On the correlation between local intrinsic dimensionality and outlierness. In Proceedings of the International Conference on Similarity Search and Applications, Lima, Peru, 7–9 October 2018; pp. 177–191.
28. Carter, K.M.; Raich, R.; Finn, W.G.; Hero, A.O., III. FINE: Fisher Information Non-parametric Embedding. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2093–2098. [CrossRef]
29. Ma, X.; Li, B.; Wang, Y.; Erfani, S.M.; Wijewickrema, S.N.R.; Schoenebeck, G.; Song, D.; Houle, M.E.; Bailey, J. Characterizing Adversarial Subspaces Using Local Intrinsic Dimensionality. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018; pp. 1–15.
30. Amsaleg, L.; Bailey, J.; Barbe, D.; Erfani, S.M.; Houle, M.E.; Nguyen, V.; Radovanović, M. The Vulnerability of Learning to Adversarial Perturbation Increases with Intrinsic Dimensionality. In Proceedings of the IEEE Workshop on Information Forensics and Security, Rennes, France, 4–7 December 2017; pp. 1–6.
31. Amsaleg, L.; Bailey, J.; Barbe, A.; Erfani, S.M.; Furon, T.; Houle, M.E.; Radovanović, M.; Nguyen, X.V. High Intrinsic Dimensionality Facilitates Adversarial Attack: Theoretical Evidence. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 854–865. [CrossRef]
32. Ma, X.; Wang, Y.; Houle, M.E.; Zhou, S.; Erfani, S.M.; Xia, S.; Wijewickrema, S.N.R.; Bailey, J. Dimensionality-Driven Learning with Noisy Labels. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 3361–3370.
33. Ansuini, A.; Laio, A.; Macke, J.H.; Zoccolan, D. Intrinsic dimension of data representations in deep neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 6111–6122.
34. Pope, P.; Zhu, C.; Abdelkader, A.; Goldblum, M.; Goldstein, T. The intrinsic dimension of images and its impact on learning. In Proceedings of the International Conference on Learning Representations, Virtual Event, 3–7 May 2021.
35. Gong, S.; Boddeti, V.N.; Jain, A.K. On the intrinsic dimensionality of image representations. In Proceedings of the CVPR, Long Beach, CA, USA, 5–20 June 2019; pp. 3987–3996.
36. Barua, S.; Ma, X.; Erfani, S.M.; Houle, M.H.; Bailey, J. Quality Evaluation of GANs Using Cross Local Intrinsic Dimensionality. *arXiv* **2019**, arXiv:1905.00643.
37. Romano, S.; Chelly, O.; Nguyen, V.; Bailey, J.; Houle, M.E. Measuring Dependency via Intrinsic Dimensionality. In Proceedings of the ICPR16, Cancun, Mexico, 4–8 December 2016; pp. 1207–1212.
38. Lucarini, V.; Faranda, D.; de Freitas, A.; de Freitas, J.; Holland, M.; Kuna, T.; Nicol, M.; Todd, M.; Vaienti, S. *Extremes and Recurrence in Dynamical Systems*; Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts; Wiley: Hoboken, NJ, USA, 2016.
39. Levina, E.; Bickel, P.J. Maximum Likelihood Estimation of Intrinsic Dimension. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 13–18 December 2004; pp. 777–784.
40. Amsaleg, L.; Chelly, O.; Furon, T.; Girard, S.; Houle, M.E.; Kawarabayashi, K.; Nett, M. Extreme-Value-Theoretic Estimation of Local Intrinsic Dimensionality. *Data Min. Knowl. Discov.* **2018**, *32*, 1768–1805. [CrossRef]
41. Hill, B.M. A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Stat.* **1975**, *3*, 1163–1174. [CrossRef]
42. Johnsson, K.; Soneson, C.; Fontes, M. Low bias local intrinsic dimension estimation from expected simplex skewness. *IEEE TPAMI* **2015**, *37*, 196–202. [CrossRef] [PubMed]
43. Amsaleg, L.; Chelly, O.; Houle, M.E.; Kawarabayashi, K.; Radovanović, R.; Treeratanajaru, W. Intrinsic dimensionality estimation within tight localities. In Proceedings of the 2019 SIAM International Conference on Data Mining, Calgary, AB, Canada, 2–4 May 2019; pp. 181–189.
44. Farahmand, A.M.; Szepesvári, C.; Audibert, J.Y. Manifold-adaptive dimension estimation. In Proceedings of the 24th International Conference on Machine Learning, Corvalis, OR, USA, 20–24 June 2007; pp. 265–272.
45. Block, A.; Jia, Z.; Polyanskiy, Y.; Rakhlin, A. Intrinsic Dimension Estimation Using Wasserstein Distances. *arXiv* **2021**, arXiv:2106.04018.
46. Thordsen, E.; Schubert, E. ABID: Angle Based Intrinsic Dimensionality—Theory and analysis. *Inf. Syst.* **2022**, *108*, 101989. [CrossRef]
47. Carter, K.M.; Raich, R.; Hero III, A.O. On Local Intrinsic Dimension Estimation and Its Applications. *IEEE Trans. Signal Process.* **2010**, *58*, 650–663. [CrossRef]

48. Tempczyk, P.; Golinski, A.; Spurek, P.; Tabor, J. LIDL: Local Intrinsic Dimension estimation using approximate Likelihood. In Proceedings of the ICLR 2021 Workshop on Geometrical and Topological Representation Learning, Online, 7 May 2021.
49. Cover, T.M.; Thomas, J.A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
50. Rioul, O. Information Theoretic Proofs of Entropy Power Inequalities. *IEEE Trans. Inf. Theory* **2011**, *57*, 33–55. [CrossRef]
51. Jelinek, F.; Mercer, R.L.; Bahl, L.R.; Baker, J.K. Perplexity—A measure of the difficulty of speech recognition tasks. *J. Acoust. Soc. Am.* **1977**, *62*, S63. [CrossRef]
52. Jost, L. Entropy and diversity. *Oikos* **2006**, *113*, 363–375. [CrossRef]
53. Kostal, L.; Lansky, P.; Pokora, O. Measures of statistical dispersion based on Shannon and Fisher information concepts. *Inf. Sci.* **2013**, *235*, 214–223. [CrossRef]
54. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control.* **1959**, *2*, 101–112. [CrossRef]
55. Di Crescenzo, A.; Longobardi, M. On cumulative entropies. *J. Stat. Plan. Inference* **2009**, *139*, 4072–4087. [CrossRef]
56. Rao, M.; Chen, Y.; Vemuri, B.C.; Wang, F. Cumulative residual entropy: A new measure of information. *IEEE Trans. Inf. Theory* **2004**, *50*, 1220–1228. [CrossRef]
57. Nguyen, H.V.; Mandros, P.; Vreeken, J. Universal Dependency Analysis. In Proceedings of the 2016 SIAM International Conference on Data Mining, Miami, FL, USA, 5–7 May 2016; pp. 792–800. [CrossRef]
58. Böhm, K.; Keller, F.; Müller, E.; Nguyen, H.V.; Vreeken, J. CMI: An Information-Theoretic Contrast Measure for Enhancing Subspace Cluster and Outlier Detection. In Proceedings of the 13th SIAM International Conference on Data Mining, Austin, TX, USA, 2–4 May 2013; pp. 198–206. [CrossRef]
59. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
60. Calì, C.; Longobardi, M.; Ahmadi, J. Some properties of cumulative Tsallis entropy. *Phys. A Stat. Mech. Its Appl.* **2017**, *486*, 1012–1021. [CrossRef]
61. Pele, D.T.; Lazar, E.; Mazurencu-Marinescu-Pele, M. Modeling Expected Shortfall Using Tail Entropy. *Entropy* **2019**, *21*, 1204. [CrossRef]
62. MacKay, D.J. *Information Theory, Inference, and Learning Algorithms*, 1st ed.; Cambridge University Press: Cambridge, UK, 2003.
63. Kac, M.; Kiefer, J.; Wolfowitz, J. On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Stat.* **1955**, *26*, 189–211. [CrossRef]
64. Nowozin, S.; Cseke, B.; Tomioka, R. f-GAN: Training generative neural samplers using variational divergence minimization. In Proceedings of the 30th Annual Conference on Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 271–279.
65. Contreras-Reyes, J. Asymptotic form of the Kullback-Leibler divergence for multivariate asymmetric heavy-tailed distributions. *Phys. A Stat. Mech. Its Appl.* **2014**, *395*, 200–208. [CrossRef]
66. Houle, M.E.; Kashima, H.; Nett, M. Generalized Expansion Dimension. In Proceedings of the IEEE 12th International Conference on Data Mining Workshops, Brussels, Belgium, 10 December 2012; pp. 587–594.
67. Karger, D.R.; Ruhl, M. Finding nearest neighbors in growth-restricted metrics. In Proceedings of the 34th ACM Symposium on Theory of Computing, Montreal, QC, Canada, 19–21 May 2002; pp. 741–750.
68. Houle, M.E. Dimensionality, Discriminability, Density and Distance Distributions. In Proceedings of the IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, USA, 7–10 December 2013; pp. 468–473.
69. Karamata, J. Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bull. Société Mathématique Fr.* **1933**, *61*, 55–62. [CrossRef]
70. Coles, S.; Bawa, J.; Trenner, L.; Dorazio, P. *An Introduction to Statistical Modeling of Extreme Values*; Springer: Berlin/Heidelberg, Germany, 2001; Volume 208.
71. Houle, M.E. Local Intrinsic Dimensionality II: Multivariate Analysis and Distributional Support. In Proceedings of the International Conference on Similarity Search and Applications, Munich, Germany, 4–6 October 2017; pp. 80–95.
72. Song, K. Renyi information, log likelihood and an intrinsic distribution measure. *J. Statist. Plann. Inference* **2001**, *93*, 51–69. [CrossRef]
73. Buono, F.; Longobardi, M. Varentropy of past lifetimes. *arXiv* **2020**, arXiv:2008.07423.
74. Maadani, S.; Borzadaran, G.R.M.; Roknabadi, A.H.R. Varentropy of order statistics and some stochastic comparisons. *Commun. Stat. Theory Methods* **2021**, *51*, 6447–6460. [CrossRef]
75. Raqab, M.Z.; Bayoud, H.A.; Qiu, G. Varentropy of inactivity time of a random variable and its related applications. *IMA J. Math. Control. Inf.* **2021**, *39*, 132–154. [CrossRef]
76. Kullback, S.; Leibler, R. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
77. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [CrossRef]
78. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimising a density power divergence. *Biometrika* **1998**, *85*, 549–559. [CrossRef]
79. Hellinger, E. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *J. Für Die Reine Und Angew. Math.* **1909**, *136*, 210–271. [CrossRef]

80. Cichocki, A.; Amari, S. Families of Alpha- Beta- and Gamma- Divergences: Flexible and Robust Measures of Similarities. *Entropy* **2010**, *12*, 1532–1568. [CrossRef]
81. Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1900**, *50*, 157–175. [CrossRef]
82. Kantorovich, L.V. Mathematical Methods of Organizing and Planning Production. *Manag. Sci.* **1939**, *6*, 366–422. [CrossRef]
83. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein Generative Adversarial Networks. In Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6–11 August 2017; Precup, D., Teh, Y.W., Eds.; PMLR: Cambridge, MA, USA, 2017; Volume 70, pp. 214–223.
84. Houle, M.E. Local Intrinsic Dimensionality III: Density and Similarity. In Proceedings of the International Conference on Similarity Search and Applications, Copenhagen, Denmark, 30 September–2 October 2020.
85. Itakura, F.; Saito, S. Analysis synthesis telephony based on the maximum likelihood method. In Proceedings of the 6th International Congress on Acoustics, Tokyo, Japan, 21–28 August 1968; pp. C17–C20.
86. Fevotte, C.; Bertin, N.; Durrieu, J. Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. *Neural Comput.* **2009**, *21*, 793–830. [CrossRef]
87. Bregman, L.M. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217. [CrossRef]
88. Nielsen, F.; Nock, R. Sided and symmetrized Bregman centroids. *IEEE Trans. Inf. Theory* **2009**, *55*, 2882–2904. [CrossRef]
89. Banerjee, A.; Merugu, S.; Dhillon, I.S.; Ghosh, J. Clustering with Bregman Divergences. *J. Mach. Learn. Res.* **2005**, *6*, 1705–1749.
90. Fang, K.W.; Kotz, S.; Wang Ng, K. *Symmetric Multivariate and Related Distributions*; CRC Press: Boca Raton, FL, USA, 2018.
91. Baker, J.A. Integration of Radial Functions. *Math. Mag.* **1999**, *72*, 392–395. [CrossRef]