Relationships between Local Intrinsic Dimensionality and Tail Entropy

James Bailey[1], Michael E. Houle[2], and Xingjun Ma[3,1]

[1] The University of Melbourne, School of Computing and Information Systems
700 Swanston St., Melbourne VIC 3010, Australia
baileyj@unimelb.edu.au
[2] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
meh@nii.ac.jp
[3] Deakin University, School of Information Technology
221 Burwood Hwy., Burwood VIC 3125, Australia
daniel.ma@deakin.edu.au

Abstract. The local intrinsic dimensionality (LID) model assesses the complexity of data within the vicinity of a query point, through the growth rate of the probability measure within an expanding neighborhood. In this paper, we show how LID is asymptotically related to the entropy of the lower tail of the distribution of distances from the query. We establish tight relationships for cumulative Shannon entropy, entropy power, and their generalized Tsallis entropy variants, all with the potential for serving as the basis for new estimators of LID, or as substitutes for LID-based characterization and feature representations in classification and other learning contexts.

1 Introduction

Assessing the complexity of high dimensional data is a fundamental task that underpins many activities in machine learning and data mining. One well-known measure of data complexity is the intrinsic dimensionality, a unitless quantity that can be interpreted as the minimum number of latent variables needed to describe the data.

The many extant formulations of intrinsic dimensionality can be divided into two broad groups, global and local. Global intrinsic dimensionality, which takes contributions from the full dataset to measure its complexity as a whole, has been more widely investigated. By contrast, local variants of intrinsic dimensionality assess the complexity of the data in the vicinity of a designated query location, most notably in terms of the growth rate in the probability measure captured by an expanding neighborhood. Local variants can therefore associate different intrinsic dimensional values to different locations in the data domain.

Our focus in this paper is on the local intrinsic dimension (LID) as formulated in [22, 23], and in particular, establishing how it relates to entropy, perhaps the most fundamental and widely-used model of data complexity. In its essence, entropy can be regarded as a measure of the uncertainty of a distribution. Our study of entropy considers the distribution of distances to a query location, where the distances are induced by a global data distribution. In particular, we consider the entropy of the lower tail of the neighbor distance distribution (the *tail entropy*), and consider its asymptotic tendency as the neighborhood radius approaches zero.

Our analysis of the relationship between the tail entropy and local intrinsic dimensionality has further implications due to an established relationship between the latter and the statistical theory of extreme values (EVT) [2]. For any distribution of distances satisfying appropriate smoothness assumptions in the lower tail, as the neighborhood radius approaches zero, the tail distribution takes the form of a power law. Asymptotically, power law distributions can be said to arise naturally in the lower tail, with the exponent of the power law corresponding to the LID value.

We formulate asymptotic results that relate local intrinsic dimensionality with multiple variants of tail entropy. In particular, we relate LID to:

- **The cumulative tail entropy**. Cumulative entropy [17, 35] 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.
- **The tail entropy power**. 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 [16], and can be interpreted as a measure of statistical dispersion [33]. It is also related to Fisher information via Stam's inequality [46].
- **Generalized tail entropies** (tail cumulative q-entropy and tail q-entropy power). Generalized Tsallis entropies [47, 8] 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 $q \to 1$. The use of such a parameter can often facilitate more accurate fitting of data characteristics and robustness to outliers.

We believe our theoretical results are interesting in their own right, as they capture fundamental properties of local neighborhood geometry, and since they hold asymptotically for essentially all smooth data distributions. The relationships between LID and tail entropy formulations also have two interesting potential applications:

- **Estimation:** Our theory allows the development of new estimators for the LID of a query point, by applying existing estimators for cumulative entropy [17] and cumulative q-entropy [8] to samples of a sufficiently-small neighborhood of the query.

- **Feature representation**: LID estimates can be used as features or as characterizations within machine learning models, such as for the detection of adversarial examples [36] or overfitting during learning [37]. However, small errors in the estimation of LID can have a disproportionally large impact on learning models. In contrast, the tail entropy power has long been known to possess attractive properties for linear discrete systems [43], and thus has potential as a more robust substitute for LID when used as a feature in logistic regression models.

In summary, our key contributions are the development of new theory that asymptotically relates tail entropy and LID, with potential applications of this theory for estimation and feature representation. To the best of our knowledge, this is the first work relating intrinsic dimensionality and the asymptotic behavior of entropy within neighborhoods of a data domain.

2 Related Work

Our work relates to intrinsic dimensionality and its estimation, as well as tail entropy and its varieties such as generalized tail entropy and cumulative tail entropy. We briefly review these in turn.

Intrinsic dimensionality can be assessed either globally (for all data points) or locally (with respect to a chosen query point). Surveys of the field provide more detail [9, 11, 48]. In the global case, considerable work has focused on topological models, with accompanying estimation methods [7, 41, 38]. Examples here include PCA and its variants [29], graph based methods [15] and fractal models [9, 20]. Other techniques such as IDEA [45, 44] and DANCo [13] estimate the dimension based on concentration of norms and angles, or 2-nearest neighbors [18].

For local intrinsic dimensionality, a popular estimator is the maximum likelihood estimator, studied in the Euclidean setting by Levina and Bickel [34] and later formulated under the more general assumptions of EVT by Amsaleg et al. [23, 2], who showed it to be equivalent to the classic Hill estimator [21]. Other local estimators include expected simplex skewness [28], the tight locality estimator [3], the MiND framework [44] and the manifold adaptive dimension [19].

Local intrinsic dimensionality has been used in a range of applications. These include modeling deformation in complex materials [49], dimension reduction via local PCA [30], similarity search [26], clustering [10], outlier detection [27], statistical manifold learning [12], adversarial example detection [36], and adversarial nearest neighbor characterization [1, 4], and deep learning understanding [37, 5]. 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 [36]. It has also been found that the LID of deep representations [5] or input data [42] is an indicator of the generalization performance of deep neural networks (DNNs). A manifold 'dimensionality expansion' phenomenon has been observed when DNNs overfit to noisy class labels [37].

Cumulative entropy was formulated in [17] and is a variant of cumulative residual entropy [35]. Outside of reliability theory analysis, it has been used in such data mining tasks as dependency analysis [39] and subspace cluster analysis [6], where it has proved effective due to the existence of good estimators. Such investigation has been at a global level (over the entire data domain), rather than at the local level as in our study. Generalized variants based on Tsallis q-statistics have been developed for both entropy [47] and cumulative entropy [8].

The concept of tail entropy has been used in financial applications for assessing the expected shortfall [40] in the upper tail using quantization. This is different from our context, where we analyze lower tails and develop exact results for an asymptotic regime.

3 Local Intrinsic Dimensionality

In this section, we summarize the LID model using the formulation of [23].

LID can be regarded as a continuous extension of the expansion dimension due to Karger and Ruhl [25, 32]. Like earlier expansion-based models of intrinsic dimension, it draws its motivation from the relationship between volume and radius in an expanding ball, where (as originally stated in [22]) 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 [23]) generalizes this notion to real-valued functions F for which $F(0) = 0$, under appropriate assumptions of smoothness.

Definition 1 ([23]). *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:*

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

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

Theorem 1 ([23]). *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, 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 has been defined. To any generated sample $\mathbf{y} \in \mathcal{D}$ we can associate the distance $r = d(\mathbf{x}, \mathbf{y})$; in this way, the global distribution that produces samples $\mathbf{y}$ can be said to induce a local distance distribution with CDF F with respect to $\mathbf{x}$. 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 [31] for the upper tails of regularly varying functions. For more information on EVT and how the LID model relates to it, we refer the reader to [14, 23, 24].

Theorem 2 (LID Representation Theorem [23]). *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*

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

whenever the integral exists.

In [23], conditions on x and w are provided for which the factor $G_F(x, w)$ 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 $G_F(x, w)$, which is related to the slowly-varying component of functions as studied in EVT [14]. 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 Tail Entropy and LID

In this section, we will establish relationships between local intrinsic dimensionality and several forms of entropy *conditioned* on the lower tails of smooth functions on domains bounded from below at zero. The results presented in this section all hold *asymptotically*, as the tail boundary tends toward zero, when *normalized* with respect to the length of the tail.

4.1 Definitions of Tail Entropy Variants

We begin with formal definitions of the tail entropies considered in this paper. In each case, we assume that we are given a non-negative real-valued function F whose restriction to $[0, w]$ satisfies the following smooth growth properties:

- $F(0) = 0$, and $F(t) > 0$ for $t \in (0, w]$;
- F is strictly monotonically increasing;
- F is continuously differentiable.

The function $\phi(t) \triangleq F(t)/F(w)$ thus satisfies the conditions of a cumulative distribution function over $t \in [0, w]$ (recall that $F(t|t \leq w) = F(t)/F(w)$ over $t \in [0, w]$), with the derivative $\phi'(t) = F'(t)/F(w)$ as its corresponding probability density function.

The following tail entropy formulations apply to any function F satisfying the conditions stated above. In their definitions, the only difference between the tail variants and the original versions is that the distribution is conditioned to the lower tail $[0, w]$. Consequently, in the tail variants, integration is performed over the lower tail and not the entire distributional range $[0, +\infty)$.

Definition 2 (Tail Entropy). *The entropy of F conditioned on $[0, w]$ is*

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

The cumulative entropy is a variant of entropy proposed in [17, 35] due to its attractive theoretical properties. Tail conditioning on the cumulative entropy has the same general form as that of the tail entropy.

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

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

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 4 (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 Tallis [47]. In general, q-entropy formulations can be shown to be identical to their Shannon entropy analogues in the limit as q tends to 1.

Table 1: Asymptotic relationships between normalized tail entropy variants and local intrinsic dimensionality.

Entropy Variant	**Normalized Tail Entropy**	**Limit as $\mathbf{w} \to \mathbf{0}^+$**
Cumulative Entropy	$\mathrm{ncH}(F,w) \triangleq \frac{1}{w}\mathrm{cH}(F,w)$	$\frac{\mathrm{ID}^*_F}{(\mathrm{ID}^*_F+1)^2}$
Cumulative q-Entropy	$\mathrm{ncH}_q(F,w) \triangleq \frac{1}{w}\mathrm{cH}_q(F,w)$	$\frac{\mathrm{ID}^*_F}{(\mathrm{ID}^*_F+1)(q\,\mathrm{ID}^*_F+1)}$
Entropy Power	$\mathrm{nHP}(F,w) \triangleq \frac{1}{w}\mathrm{HP}(F,w)$	$\frac{1}{\mathrm{ID}^*_F}\exp\left(1-\frac{1}{\mathrm{ID}^*_F}\right)$
q-Entropy Power	$\mathrm{nHP}_q(F,w) \triangleq \frac{1}{w}\mathrm{HP}_q(F,w)$	$\left(\frac{(\mathrm{ID}^*_F)^q}{q\,\mathrm{ID}^*_F-q+1}\right)^{\frac{1}{1-q}}$

Definition 5 (Tail q-Entropy). *For any $q > 0$ ($q \neq 1$), the q-entropy of F conditioned on $[0, w]$ is defined to be*

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

Definition 6 (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*

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

We define the tail q-entropy power using the q-exponential function from Tsallis statistics [47], $\exp_q(x) \triangleq [1+(1-q)x]^{\frac{1}{1-q}}$. Note that L'Hôpital's rule can be used to show that $\exp_q(x) \to e^x$ as $q \to 1$.

Definition 7 (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*

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

For the cumulative tail entropy and tail entropy power variants, we will also consider a normalization given by the ratio of the entropy with w, the length of the tail. In the remainder of this section, we will show that as w tends to zero, the limits of these normalized entropies can be expressed in terms of the local intrinsic dimensionality of F. The notation for these normalized entropy variants, and our theorems for their limits in terms of LID, are summarized in Table 1.

4.2 Technical Preliminaries

Before presenting the main theoretical results of the paper, we begin with two technical lemmas. The first lemma concerns a slight generalization of the cumulative entropy formulation, that allows it to greatly facilitate the proofs for two tail entropy variants, the cumulative entropy and the entropy power.

Lemma 1. *Let $F : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ be a function such that $F(0) = 0$, and assume that ID^*_F exists and is positive. For some value of $r > 0$, let us further assume that within the interval $[0, r)$, F is continuously differentiable and strictly monotonically increasing, and ψ is positive. Then for any constant $u < \mathrm{ID}^*_F$,*

$$\lim_{w\to 0^+} w^{u-1} \int_0^w \frac{\psi(w)\,F(t)}{t^u F(w)} \ln \frac{\psi(w)\,F(t)}{t^u F(w)}\,\mathrm{d}t = \lim_{w\to 0^+} \frac{\psi(w)}{\mathrm{ID}^*_F + 1 - u} \left[\ln \frac{\psi(w)}{w^u} - \frac{\mathrm{ID}^*_F - u}{\mathrm{ID}^*_F + 1 - u} \right]$$

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

Proof: Since the limit $\mathrm{ID}^*_F = \lim_{v\to 0^+} \mathrm{ID}_F(v)$ is assumed to exist, we have that for any real value $\epsilon > 0$ satisfying $\epsilon < \min\{r, \mathrm{ID}^*_F - u\}$, there must exist a value $0 < \delta < \epsilon$ such that $v < \delta$ implies that $|\,\mathrm{ID}_F(v) - \mathrm{ID}^*_F\,| < \epsilon$. Therefore, when $0 < t \leq w < \delta$,

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

Exponentiating, we obtain the bounds

$$\left(\frac{w}{t}\right)^{-\epsilon} \;<\; G_F\,(t, w) \;<\; \left(\frac{w}{t}\right)^{\epsilon}. \tag{1}$$

For any real $x > 0$, we define $\mathrm{xlnx}(x) \triangleq x \ln x$. Applying Theorem 2 to $F(t)$, and making use of the upper bound on G_F, the integral becomes

$$\begin{aligned} &\int_0^w \mathrm{xlnx}\left(\frac{\psi(w)\,F(t)}{t^u F(w)}\right)\mathrm{d}t \\ &= \int_0^w \mathrm{xlnx}\left(\frac{\psi(w)}{t^u}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F} G_F(t, w)\right)\mathrm{d}t < \int_0^w \mathrm{xlnx}\left(\frac{\psi(w)}{t^u}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F}\left(\frac{w}{t}\right)^{\epsilon}\right)\mathrm{d}t \\ &< \int_0^w \mathrm{xlnx}\left(\frac{\psi(w)}{t^u}\left(\frac{t}{w}\right)^{\mathrm{ID}^*_F - \epsilon}\right)\mathrm{d}t < \frac{\psi(w)}{w^{m+u}} \int_0^w t^m \cdot \left[m \ln t + \ln \frac{\psi(w)}{w^{m+u}} \right] \mathrm{d}t, \end{aligned} \tag{2}$$

where $m \triangleq \mathrm{ID}^*_F - u - \epsilon > 0$.

Noting that $m > 0$ implies that $\lim_{t\to 0} t^m \ln t = 0$, integration of Equation 2 by parts yields an expression that depends on F only through its LID value.

$$\begin{aligned} w^{u-1} &\int_0^w \mathrm{xlnx}\left(\frac{\psi(w)\,F(t)}{t^u F(w)}\right)\mathrm{d}t \\ &< \frac{m w^{u-1} \psi(w)}{w^{m+u}} \left[\frac{t^{m+1}}{m+1} \ln t \,\bigg|_0^w - \int_0^w \frac{t^{m+1}}{m+1} \cdot \frac{1}{t}\,\mathrm{d}t \right] \\ &\quad + \frac{w^{u-1}\psi(w)}{w^{m+u}} \ln \frac{\psi(w)}{w^{m+u}} \cdot \frac{w^{m+1}}{m+1} \end{aligned}$$

$$
\begin{aligned}
&< \frac{m\psi(w)}{w^{m+1}}\left[\frac{w^{m+1}}{m+1}\ln w - \frac{w^{m+1}}{(m+1)^2}\right] + \frac{\psi(w)}{m+1}\ln\frac{\psi(w)}{w^{m+u}} \\
&< \frac{\psi(w)}{m+1}\left[m\ln w - \frac{m}{m+1} + \ln\frac{\psi(w)}{w^{m+u}}\right] \;<\; \frac{\psi(w)}{m+1}\left[\ln\frac{\psi(w)}{w^{u}} - \frac{m}{m+1}\right] \\
&= \frac{\psi(w)}{\mathrm{ID}^*_F + 1 - u - \epsilon}\left[\ln\frac{\psi(w)}{w^{u}} - \frac{\mathrm{ID}^*_F - u - \epsilon}{\mathrm{ID}^*_F + 1 - u - \epsilon}\right] .
\end{aligned}
$$

Similar arguments using the lower bound from Equation 1 leads us to

$$
w^{u-1}\int_0^w \mathrm{xlnx}\left(\frac{\psi(w)\,F(t)}{t^u F(w)}\right)\mathrm{d}t \;>\; \frac{\psi(w)}{\mathrm{ID}^*_F + 1 - u + \epsilon}\left[\ln\frac{\psi(w)}{w^{u}} - \frac{\mathrm{ID}^*_F - u + \epsilon}{\mathrm{ID}^*_F + 1 - u + \epsilon}\right] .
$$

Since ϵ can be chosen arbitrarily close to 0, and since $0 < w < \epsilon$ by construction, taking the limit as $w \to 0^+$ yields

$$
\lim_{w\to 0^+} w^{u-1}\int_0^w \mathrm{xlnx}\left(\frac{\psi(w)\,F(t)}{t^u F(w)}\right)\mathrm{d}t \;=\; \lim_{w\to 0^+} \frac{\psi(w)}{\mathrm{ID}^*_F + 1 - u}\left[\ln\frac{\psi(w)}{w^{u}} - \frac{\mathrm{ID}^*_F - u}{\mathrm{ID}^*_F + 1 - u}\right]
$$

whenever the right-hand limit exists, or diverges to $+\infty$ or $-\infty$. □

The second technical lemma follows as a corollary of Lemma 1, since it uses much of the same proof strategy, albeit more simply and directly. Analogous with Lemma 1, it concerns a slight generalization of the cumulative q-entropy formulation that facilitates the proof of the results for the q-entropy and q-entropy power variants.

Corollary 1. *Let $F : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ be a function such that $F(0) = 0$, and assume that ID^*_F exists and is positive. For some value of $r > 0$, let us further assume that within the interval $[0, r)$, F is continuously differentiable and strictly monotonically increasing, and ψ is positive. Then for any constants $u < \mathrm{ID}^*_F$ and $z > 0$,*

$$
\lim_{w\to 0^+} w^{zu-1}\int_0^w \left(\frac{\psi(w)\,F(t)}{t^u F(w)}\right)^z \mathrm{d}t \;=\; \frac{\lim_{w\to 0^+}\psi^z(w)}{z\,\mathrm{ID}^*_F - zu + 1}
$$

whenever the right-hand limit exists, or diverges to $+\infty$ or $-\infty$.

Proof: Following the same proof strategy of Lemma 1 that led to Equation 2, we arrive at the following upper bound on the integral:

$$
\int_0^w \left(\frac{\psi(w)\,F(t)}{t^u F(w)}\right)^z \mathrm{d}t \;<\; \frac{\psi^z(w)}{w^{z(m+u)}}\int_0^w t^{zm}\,\mathrm{d}t \;=\; \frac{\psi^z(w)}{(zm+1)w^{zu-1}} ,
$$

where $m = \mathrm{ID}^*_F - u - \epsilon$ as before.

Continuing according to the proof strategy of Lemma 1, we use the lower bound from Equation 1, let ϵ vanish, and then apply the limit $w \to 0^+$ with a factor of w^{zu-1}. This brings us to

$$\lim_{w\to 0^+} w^{zu-1} \int_0^w \left(\frac{\psi(w)\, F(t)}{t^u F(w)} \right)^z \mathrm{d}t$$
$$= \lim_{w\to 0^+} w^{zu-1} \frac{\psi^z(w)}{(z\,\mathrm{ID}^*_F - zu + 1) w^{zu-1}} \;=\; \frac{\lim_{w\to 0^+} \psi^z(w)}{z\,\mathrm{ID}^*_F - zu + 1}\,,$$

as required. □

4.3 Cumulative Tail Entropy and LID

Using the technical lemmas established in Section 4.2, we present the main results for the cumulative tail entropy variants. The first result shows that as the tail length w tends to zero, the normalized cumulative entropy $\mathrm{ncH}(F,w) \triangleq \frac{1}{w}\mathrm{cH}(F,w)$ tends to a value entirely determined by the local intrinsic dimensionality associated with F.

Theorem 3. *Let $F : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ be a function such that $F(0) = 0$, and assume that ID^*_F exists and is positive. For some value of $r > 0$, let us further assume that within the interval $[0, r)$, F is continuously differentiable and strictly monotonically increasing. We have*

$$\lim_{w\to 0^+} \mathrm{ncH}(F,w) \;=\; \lim_{w\to 0^+} -\frac{1}{w} \int_0^w \frac{F(t)}{F(w)} \ln \frac{F(t)}{F(w)}\, \mathrm{d}t \;=\; \frac{\mathrm{ID}^*_F}{(\mathrm{ID}^*_F + 1)^2}\,.$$

Proof: Follows directly from Lemma 1, for the choices $u = 0$ and $\psi(w) = 1$. □

The second result uses Corollary 1 to show that as the tail length w tends to zero, the normalized cumulative q-entropy $\mathrm{ncH}_q(F,w) \triangleq \frac{1}{w}\mathrm{cH}_q(F,w)$ tends to a value determined by q together with the local intrinsic dimensionality associated with F.

Theorem 4. *Let $F : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ be a function such that $F(0) = 0$, and assume that ID^*_F exists and is positive. For some value of $r > 0$, let us further assume that within the interval $[0, r)$, F is continuously differentiable and strictly monotonically increasing. Then for $q > 0$ with $q \neq 1$,*

$$\lim_{w\to 0^+} \mathrm{ncH}_q(F,w)$$
$$= \lim_{w\to 0^+} \frac{1}{w(q-1)} \int_0^w \frac{F(t)}{F(w)} - \left(\frac{F(t)}{F(w)} \right)^q \mathrm{d}t \;=\; \frac{\mathrm{ID}^*_F}{(\mathrm{ID}^*_F + 1)(q\,\mathrm{ID}^*_F + 1)}\,.$$

Proof: Separating the integral and applying Corollary 1 twice,

$$\lim_{w\to 0^+} \frac{1}{w(q-1)} \int_0^w \frac{F(t)}{F(w)} - \left(\frac{F(t)}{F(w)} \right)^q \mathrm{d}t$$
$$= \frac{1}{q-1} \left(\frac{1}{\mathrm{ID}^*_F + 1} - \frac{1}{q\,\mathrm{ID}^*_F + 1} \right) \;=\; \frac{\mathrm{ID}^*_F}{(\mathrm{ID}^*_F + 1)(q\,\mathrm{ID}^*_F + 1)}$$

follows for the choices $u = 0$, $\psi(w) = 1$, and (respectively) $z = 1$ and $z = q$. □

Observe that as q tends to 1, the cumulative q-entropy variant $\mathrm{ncH}_q(F,w)$ does tend to the cumulative entropy $\mathrm{ncH}(F,w)$, as one would expect.

4.4 Tail Entropy Power and LID

We find that we encounter convergence issues when attempting to use the machinery of Lemma 1 to formulate a relationship between LID and either the tail entropy $\mathrm{H}(F,w)$ or the normalized tail entropy $\mathrm{nH}(F,w)$, the limits diverging as the tail size tends to zero.

Instead, we show that the entropy power, when normalized, does have a limit expressed as a function of the LID of F.

Theorem 5. *Let $F:\mathbb{R}^{\geq 0}\to\mathbb{R}^{\geq 0}$ be a function such that $F(0)=0$, and assume that ID^*_F exists and is greater than 1. For some value of $r>0$, let us further assume that within the interval $[0,r)$, F is continuously differentiable and strictly monotonically increasing. Then*

$$\begin{aligned}&\lim_{w\to 0^+}\mathrm{nHP}(F,w)\\ &\quad=\lim_{w\to 0^+}\frac{1}{w}\exp\left(-\int_0^w\frac{F'(t)}{F(w)}\ln\frac{F'(t)}{F(w)}\,\mathrm{d}t\right)\;=\;\frac{1}{\mathrm{ID}^*_F}\exp\left(1-\frac{1}{\mathrm{ID}^*_F}\right).\end{aligned}$$

Proof: Due to space limitations, the details are omitted in this version. The proof is analogous to that of Theorem 3, and makes use of Theorem 1 and Lemma 1 with the choices $u=1$ and $\psi(w)=\mathrm{ID}^*_F$. The choice of u is valid for Lemma 1 since by assumption $\mathrm{ID}^*_F>1=u$. □

For the case of the normalized tail q-entropy power $\mathrm{nHP}_q(F,w)$, we have the following result.

Theorem 6. *Let $F:\mathbb{R}^{\geq 0}\to\mathbb{R}^{\geq 0}$ be a function such that $F(0)=0$, and assume that ID^*_F exists and is greater than 1. For some value of $r>0$, let us further assume that within the interval $[0,r)$, F is continuously differentiable and strictly monotonically increasing. Then for $q>0$ $(q\neq 1)$,*

$$\begin{aligned}&\lim_{w\to 0^+}\mathrm{nHP}_q(F,w)\\ &\quad=\lim_{w\to 0^+}\frac{1}{w}\exp_q\left(\frac{1}{q-1}\left[1-\int_0^w\left(\frac{F'(t)}{F(w)}\right)^q\mathrm{d}t\right]\right)\;=\;\left[\frac{(\mathrm{ID}^*_F)^q}{q\,\mathrm{ID}^*_F-q+1}\right]^{\frac{1}{1-q}}.\end{aligned}$$

Proof: Due to space limitations, the details are omitted in this version. The proof is analogous to that of Theorem 4, and makes use of Theorem 1 and Corollary 1 with the choices $u=1$, $\psi(w)=\mathrm{ID}^*_F$, and $z=q$. The choice of u is valid for Corollary 1 since by assumption $\mathrm{ID}^*_F>1=u$. □

5 Conclusion

In this preliminary theoretical investigation, we have established an asymptotic relationship between tail entropy variants and the emerging theory of local intrinsic dimensionality. Our results provide new insights into the complexity of data within local neighborhoods, and how they may be assessed. These fundamental discoveries also open the door to cross-fertilization between intrinsic dimensionality research and entropy research, particularly as regards the potential for the use of robust estimators of tail entropy as substitutes for LID in learning contexts. Our results could also allow for applications and characterizations for DNNs based on LID to be extended to the field of information theory.

As future work, we plan to follow with in-depth experimental studies on the performance characteristics of cumulative entropy and entropy power as estimators or substitutes of LID for deep learning and data mining applications. We also plan to investigate the generalization and learning behaviors of DNNs based on both LID and tail entropy.

Acknowledgments

Michael E. Houle acknowledges the financial support of JSPS Kakenhi Kiban (B) Research Grant 18H03296.

References

1. L. Amsaleg, J. Bailey, D. Barbe, S. M. Erfani, M. E. Houle, V. Nguyen, and M. Radovanović. The vulnerability of learning to adversarial perturbation increases with intrinsic dimensionality. In *IEEE Workshop on Information Forensics and Security*, pages 1–6, 2017.
2. L. Amsaleg, O. Chelly, T. Furon, S. Girard, M. E. Houle, K. Kawarabayashi, and M. Nett. Extreme-value-theoretic estimation of local intrinsic dimensionality. *Data Mining and Knowledge Discovery*, 32(6):1768–1805, 2018.
3. L. Amsaleg, O. Chelly, M. E Houle, K. Kawarabayashi, R. Radovanović, and W. Treeratanajaru. Intrinsic dimensionality estimation within tight localities. In *Proc. 2019 SIAM International Conference on Data Mining*, page 181–189, 2019.
4. Laurent Amsaleg, James Bailey, Amélie Barbe, Sarah M. Erfani, Teddy Furon, Michael E. Houle, Miloš Radovanović, and Xuan Vinh Nguyen. High intrinsic dimensionality facilitates adversarial attack: Theoretical evidence. *IEEE Transactions on Information Forensics and Security*, 16:854–865, 2021.
5. A. Ansuini, A. Laio, J. H. Macke, and D. Zoccolan. Intrinsic dimension of data representations in deep neural networks. In *Advances in Neural Information Processing Systems*, pages 6111–6122, 2019.
6. K. Böhm, F. Keller, E. Müller, H. V. Nguyen, and J. Vreeken. CMI: an information-theoretic contrast measure for enhancing subspace cluster and outlier detection. In *Proc. 13th SIAM International Conference on Data Mining*, pages 198–206, 2013.
7. J. Bruske and G. Sommer. Intrinsic dimensionality estimation with optimally topology preserving maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(5):572–575, 1998.
8. C. Calì, M. Longobardi, and J. Ahmadi. Some properties of cumulative Tsallis entropy. *Physica A: Statistical Mechanics and its Applications*, 486:1012–1021, 2017.
9. F. Camastra and A. Staiano. Intrinsic dimension estimation: Advances and open problems. *Information Sciences*, 328:26–41, 2016.

10. P. Campadelli, E. Casiraghi, C. Ceruti, G. Lombardi, and A. Rozza. Local intrinsic dimensionality based features for clustering. In *International Conference on Image Analysis and Processing*, pages 41–50, 2013.
11. P. Campadelli, E. Casiraghi, C. Ceruti, and A. Rozza. Intrinsic dimension estimation: Relevant techniques and a benchmark framework. *Mathematical Problems in Engineering*, 2015.
12. K. M. Carter, R. Raich, W. G. Finn, and A. O. Hero III. FINE: Fisher information non-parametric embedding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2093–2098, 2009.
13. C. Ceruti, S. Bassis, A. Rozza, G. Lombardi, E. Casiraghi, , and P. Campadelli. DANCo: An intrinsic dimensionality estimator exploiting angle and norm concentration. *Pattern Recognition*, 47:2569–2581, 2014.
14. S. Coles, J. Bawa, L. Trenner, and P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.
15. J. A. Costa and A. O. Hero III. Entropic graphs for manifold learning. In *The 37th Asilomar Conference on Signals, Systems & Computers*, volume 1, pages 316–320, 2003.
16. T. M. Cover and J. A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.
17. A. Di Crescenzo and M. Longobardi. On cumulative entropies. *Journal of Statistical Planning and Inference*, 139(12):4072–4087, 2009.
18. E. Facco, M. d'Errico, A. Rodriguez, and A. Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific Reports*, 7(12140), 2017.
19. A. M. Farahmand, C. Szepesvári, and J. Y. Audibert. Manifold-adaptive dimension estimation. In *Proc. 24th International Conference on Machine Learning*, pages 265–272, 2007.
20. M. Hein and J. Y. Audibert. Intrinsic dimensionality estimation of submanifolds in R^d. In *Proc. 22nd International Conference on Machine Learning*, pages 289–296, 2005.
21. B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.
22. M. E. Houle. Dimensionality, discriminability, density and distance distributions. In *IEEE 13th International Conference on Data Mining Workshops*, pages 468–473, 2013.
23. M. E. Houle. Local intrinsic dimensionality I: an extreme-value-theoretic foundation for similarity applications. In *International Conference on Similarity Search and Applications*, pages 64–79, 2017.
24. M. E. Houle. Local intrinsic dimensionality II: multivariate analysis and distributional support. In *International Conference on Similarity Search and Applications*, pages 80–95, 2017.
25. M. E. Houle, H. Kashima, and M. Nett. Generalized expansion dimension. In *IEEE 12th International Conference on Data Mining Workshops*, pages 587–594, 2012.
26. M. E. Houle, X. Ma, M. Nett, and V. Oria. Dimensional testing for multi-step similarity search. In *IEEE 12th International Conference on Data Mining*, pages 299–308, 2012.
27. M. E. Houle, E. Schubert, and A. Zimek. On the correlation between local intrinsic dimensionality and outlierness. In *International Conference on Similarity Search and Applications*, page 177–191, 2018.
28. K. Johnsson, C. Soneson, and M. Fontes. Low bias local intrinsic dimension estimation from expected simplex skewness. *IEEE TPAMI*, 37(1):196–202, 2015.

29. I. T. Jolliffe. *Principal Component Analysis.* Springer, 2002.
30. N. Kambhatla and T. K. Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 9(7):1493–1516, 1997.
31. J. Karamata. Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bulletin de la Société Mathématique de France*, 61:55–62, 1933.
32. D. R. Karger and M. Ruhl. Finding nearest neighbors in growth-restricted metrics. In *Proc. 34th ACM Symposium on Theory of Computing*, pages 741–750, 2002.
33. L. Kostal, P. Lansky, and O. Pokora. Measures of statistical dispersion based on Shannon and Fisher information concepts. *Information Sciences*, 2013.
34. E. Levina and P. J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, pages 777–784, 2004.
35. M. Rao, Y. Chen, B. C. Vemuri, and F. Wang. Cumulative residual entropy: a new measure of information. *IEEE Transactions on Information Theory*, 50(6):1220–1228, 2004.
36. X. Ma, B. Li, Y. Wang, S. M. Erfani, S. N. R. Wijewickrema, G. Schoenebeck, D. Song, M. E. Houle, and J. Bailey. Characterizing adversarial subspaces using local intrinsic dimensionality. In *International Conference on Learning Representations*, pages 1–15, 2018.
37. X. Ma, Y. Wang, M. E. Houle, S. Zhou, S. M. Erfani, S.-T. Xia, S. N. R. Wijewickrema, and J. Bailey. Dimensionality-driven learning with noisy labels. In *International Conference on Machine Learning*, pages 3361–3370, 2018.
38. G. Navarro, R. Paredes, N. Reyes, and C. Bustos. An empirical evaluation of intrinsic dimension estimators. *Inf. Syst.*, 64:206–218, 2017.
39. H. V. Nguyen, P. Mandros, and J. Vreeken. Universal dependency analysis. In *Proc. 2016 SIAM International Conference on Data Mining*, pages 792–800, 2016.
40. D. T. Pele, E. Lazar, and M. Mazurencu-Marinescu-Pele. Modeling expected shortfall using tail entropy. *Entropy*, 21(12):1204, 2019.
41. K. W. Pettis, Bailey. T. A., A. K. Jain, and R. C. Dubes. An intrinsic dimensionality estimator from near-neighbor information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:25—37, 1979.
42. P. Pope, C. Zhu, A. Abdelkader, M. Goldblum, and T. Goldstein. The intrinsic dimension of images and its impact on learning. In *International Conference on Learning Representations*, 2021.
43. H. C. Ratz. Entropy power factors for linear discrete systems. *Canadian Electrical Engineering Journal*, 8(2):73–78, 1983.
44. A. Rozza, G. Lombardi, C. Ceruti, E. Casiraghi, and P. Campadelli. Novel high intrinsic dimensionality estimators. *Machine Learning*, 89(1–2):37–65, 2012.
45. A. Rozza, G. Lombardi, M. Rosa, E. Casiraghi, and P. Campadelli. IDEA: Intrinsic dimension estimation algorithm. In *International Conference on Image Analysis and Processing*, pages 433–442, 2011.
46. A. J. Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2:101–112, 1959.
47. C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 07 1988.
48. P. J. Verveer and R. P. W. Duin. An evaluation of intrinsic dimensionality estimators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):81–86, 1995.
49. S. Zhou, A. Tordesillas, M. Pouragha, J. Bailey, and H. Bondell. On local intrinsic dimensionality of deformation in complex materials. *Nature Scientific Reports*, 11(10216), 2021.