



Local Intrinsic Dimensionality III: Density and Similarity

Michael E. Houle^(✉)

National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
meh@nii.ac.jp

Abstract. In artificial intelligence, machine learning, and other areas in which statistical estimation and modeling is common, distributions are typically assumed to admit a representation in terms of a probability density function (pdf). However, in many situations, such as mixture modeling and subspace methods, the distributions in question are not always describable in terms of a single pdf. In this paper, we present a theoretical foundation for the modeling of density ratios in terms of the local intrinsic dimensionality (LID) model, in a way that avoids the use of traditional probability density functions. These formulations provide greater flexibility when modeling data under the assumption of local variation in intrinsic dimensionality, in that no explicit dependence on a fixed-dimensional data representation is required.

1 Introduction

1.1 Probability Density and Dimensionality

Modelers in computer science disciplines such as artificial intelligence, machine learning, and pattern recognition typically base their analyses on a distributional view of their data sources, either in terms of parametric approaches involving standard distributions, or through nonparametric approaches.

For applications in which the data is continuous in nature, and in the supporting domains of continuous statistical modeling and probability theory, the concept of probability density is ubiquitous. In continuous domains, the absolute probability associated with any single event is typically zero, due to the infinite number of possible outcomes. For this reason, theoreticians and practitioners alike have been concerned with the relative likelihood of generation of one given sample value compared to another, as taken over infinitesimally-small volumes of positive probability measure that contain the values of interest. These density values, when integrated over the domain, account for its full probability measure, which is 1 by definition.

Not all distributions of interest admit a probability density function (pdf). Perhaps the most famous and important example is that of mixture distributions derived from a collection of random variables. Sampling from a mixture distribution involves a two-stage process, in which a distribution is first selected from

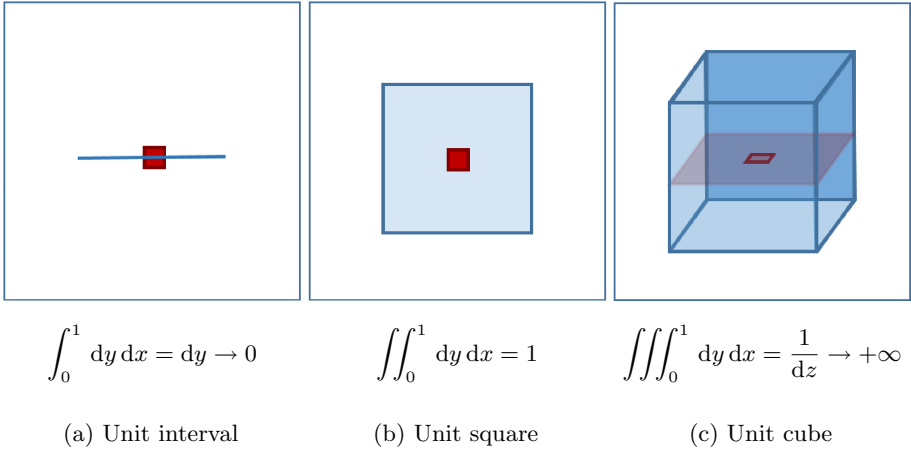


Fig. 1. Application of a 2-D pdf in uniform distributions in 1-D, 2-D, and 3-D domains. A 2-D pdf expresses probability density as a ratio between probability mass and area. Since the 1-D interval has area 0, and since the 3-D cube has infinite area, integration of the 2-D pdf over the interval and the cube (in terms of the infinitesimal product $dy dx$) would produce the values 0 and $+\infty$, respectively.

the mixture according to some probability associated with it, and then a value is generated from the selected distribution. Although each of the constituent distributions of the mixture can be associated with a probability density function, in general the mixture distribution as a whole may not. As an illustrative example, consider the situation in Fig. 1 in which with equal probability a sample can be selected from one of three uniform distributions: over a 1-dimensional unit interval, over a 2-dimensional unit square, or over a 3-dimensional unit cube. Each of the three pdfs of the mixture, when integrated over its own domain (interval, square, or cube), would return a value of 1. However, taking the 2-D pdf of the square domain and integrating it over the 1-D interval domain would produce a result of 0, whereas its integration over the 3-D cube would diverge to $+\infty$. Accordingly, it is not possible to devise a single probability density function capable of representing this mixture distribution as a whole.

Standard techniques for the direct estimation of probability density (such as multivariate kernel density estimation [35]) can be problematic when the data dimensionality is assumed to be large and fixed. As the example in Fig. 1 shows, standard formulations of probability density as an integrand over the distribution domain explicitly depend on the dimensionalities of the infinitesimal volumes over which the integration is performed. In settings for which the numbers of features is very large, data analysis models that rely on forms of density estimation can exhibit bias in terms of the number of features that are relevant to individual localities within the domain, and the degree of correlation and other interactions among these features. Variations in the numbers of relevant features from locality to locality within the domain, together with vast

differences between the (local) relevant feature dimension and the total (global) feature dimension, can greatly degrade the effectiveness of the distributional model in data analysis applications.

1.2 Volume, Dimensionality, and Similarity

Although direct estimation of probability density can be problematic, data analysis that relies on a comparison between two probability densities can (explicitly or implicitly) benefit from modeling in terms of their ratios. In recent years, density ratio estimation has been well-studied as an important area in its own right [34], and has been widely adopted throughout AI and machine learning. In addition to being used to compare the relative concentration of data between two locations within a common distribution, it has also served as the foundation of popular measures of the similarity between two data distributions on compatible domains, such as the Kullback-Leibler (KL) divergence [27] and other f -divergences [28]. Although direct estimation of ratios has been shown to be inherently biased for small sample sizes [9, 33], division of one probability density by another has the advantage of producing a unitless quantity independent of local dimensional assumptions. Nevertheless, it should be noted that KL-divergence and other distributional similarity measures typically integrate these ratios of probability density over the entire domain (of one of the two distributions), thereby reintroducing fixed-dimensional infinitesimal volumes into the data model.

The domains of continuous distributions are typically equipped with a similarity (or dissimilarity) measure through which the interrelationships among data are modeled and assessed. In many situations in machine learning/and data mining, modelers and practitioners tend to view similarity-based density estimation as an acceptable proxy for volume-based density estimation. However, in general this is not the case: volume tends to increase not linearly with distance radius, but with that radius raised to the power of the data dimension.

In this paper, we explicitly acknowledge this disparity between radius and volume by investigating the issue of density ratios from the perspective of the recent Local Intrinsic Dimensionality (LID) [15, 16] model. LID is a distributional form of intrinsic dimensional modeling in which the volume of a ball of radius r is taken to be the probability measure associated with its interior, denoted by $F(r)$. The function F can be regarded as the cumulative distribution function (cdf) of an underlying distribution of distances.

Theoretical properties of LID in multivariate analysis have been studied recently [17]. LID and related expansion-based measures of intrinsic dimensionality [18] have also seen applications in such areas as similarity search [3, 8, 19–22], dependency analysis [31], feature selection and ranking [4, 13, 23], outlier detection and its analysis [11, 24, 32], and deep learning [29, 30]. Practical estimators of LID have been developed based on interpoint distances within neighborhood samples [1, 2], which trace their roots to the estimation of the generalized pareto distribution (GPD) shape parameter [14] developed within the statistical discipline of extreme value theory (EVT).

1.3 Overview

The goal of the paper is to introduce a theoretical foundation that can serve for modeling density ratios in terms of LID, without recourse to the traditional formulation of probability density. After first briefly surveying of the LID model in the next section, in Sect. 3 we present several results relating the limits of functions with properties appropriate to density ratio modeling: namely, the functions are assumed to be smooth (continuously differentiable), positive, vanishing at zero, and of the form of a cdf of an induced distance distribution relative to a reference location in a distribution of interest. In each theoretical statement, the local intrinsic dimensionality is shown to influence the density ratio model in a natural way. The paper is then concluded in Sect. 4.

2 Local Intrinsic Dimensionality

In this section, we give a brief overview of the extreme-value-theoretic LID model of intrinsic dimensionality first introduced in [15]. For more information on the model and its connections to the statistical theory of extreme values, please see [10, 16, 17].

The LID model falls into the expansion family of intrinsic dimensional estimation [18, 26]. Like earlier expansion models, LID draws its motivation from the relationship between volume and radius in an expanding ball around points of interest. Unlike these models, the LID interprets volume as a function of the same form as a univariate cumulative distribution function (cdf), representing the probability measure $F(r)$ captured by a ball of radius r . Although motivated by applications involving the distribution of distance values, the model formulation (as stated in [16]) generalizes this notion even further, to any smooth real-valued function that is non-zero in the vicinity of $r \neq 0$.

Definition 1 ([16]). *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:*

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

Under the same assumptions on F , when F can be interpreted as the cdf of a distance distribution, the definition of LID can be regarded as equivalent to a notion of indiscriminability. Intuitively, if an underlying distance measure is indiscriminative at a given distance r , then expanding the distance by some small factor should incur a relatively large increase in probability measure as a proportion of the current value, $F(r)$. Accordingly, the indiscriminability of the distance variable is defined as the limit of the ratio of two quantities: the proportional rate of increase of $F(r)$, and the proportional rate of increase in r .

Definition 2 ([16]). *Let F be a real-valued function that is non-zero over some open interval containing $r \in \mathbb{R}$, $r \neq 0$. The indiscriminability of F at r is defined as follows whenever the limit exists:*

$$\text{InDiscr}_F(r) \triangleq \lim_{\epsilon \rightarrow 0} \left[\frac{F((1+\epsilon)r) - F(r)}{F(r)} \right] / \frac{(1+\epsilon)r - r}{r} = \lim_{\epsilon \rightarrow 0} \frac{F((1+\epsilon)r) - F(r)}{\epsilon \cdot F(r)}.$$

When F satisfies certain smoothness conditions in the vicinity of r , its intrinsic dimensionality and indiscriminability have been shown to be identical:

Theorem 1 ([16]). *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*

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

Let \mathbf{x} be any reference location within a data domain \mathcal{S} equipped with a distance measure d . To any point $\mathbf{y} \in \mathcal{D}$ we can associate the distance $r = d(\mathbf{x}, \mathbf{y})$; in this way, any global data distribution over \mathcal{D} induces a local distance distribution with respect to \mathbf{x} . Motivated by a need to characterize the local intrinsic dimensionality in the vicinity of individual reference points, we are interested in the limit of $\text{ID}_F(r)$ as the distance r tends to 0. For convenience, for non-zero distances r we refer to $\text{ID}_F(r)$ as the *indiscriminability of F at r* , and to $\text{ID}_F^* \triangleq \lim_{r \rightarrow 0} \text{ID}_F(r)$ as the *local intrinsic dimension* (or *LID*) of F .

In the ideal case where the data in the vicinity of \mathbf{x} is distributed uniformly within a submanifold in \mathcal{D} , ID_F^* would equal the dimension of the submanifold; however, in general these distributions are not ideal, the manifold model of data does not perfectly apply, and ID_F^* is not necessarily an integer. Nevertheless, the local intrinsic dimensionality would give an indication of the dimension of the submanifold containing \mathbf{x} that would best fit the data distribution in the vicinity of \mathbf{x} .

The indiscriminability function ID_F can be seen to fully characterize its associated function F .

Theorem 2 (LID Representation Theorem [16]). *Let $F : \mathbb{R} \rightarrow \mathbb{R}$ 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*

$$\frac{F(x)}{F(w)} = \left(\frac{x}{w}\right)^{\text{ID}_F^*} \cdot G_F(x, w), \text{ where}$$

$$G_F(x, w) \triangleq \exp\left(\int_x^w \frac{\text{ID}_F^* - \text{ID}_F(t)}{t} dt\right),$$

whenever the integral exists.

In [16], conditions on x and w are provided for which the $G_F(x, w)$ can be seen to tend to 1 as $x, w \rightarrow 0$. The function $G_F(x, w)$ is related to the slowly-varying

functions of long-standing interest within the EVT research community [10, 25], as it governs the rate of convergence of F to its ideal (asymptotic) form. Next, we revisit this issue so as to prove statements useful for the analysis of limits of density ratios, that do not explicitly rely on formulations involving G_F .

3 LID-Aware Density Ratios

In this section, we present statements concerning the limits of ratios of two functions, with properties that allow them to be applied to situations in which density ratios are to be modeled. Instead of considering the ratio of pdfs over an infinitesimal volume whose dimensions depend on knowledge of a full-dimensional coordinate system (as in the usual sense of density ratio estimation), we instead consider as the underlying volume spheres with infinitesimal radii. Local density ratios can be then be described as ratio of two cdf functions of distance distributions, each representing the probability measure captured by a sphere, whose radii (the function arguments) both tend to zero in some fashion.

Although this work is motivated by an interest in modeling limits of density ratios, the theoretical statements will be presented in a more generally applicable way. Here, we assume only that the functions are smooth (continuously differentiable), positive, and vanish at zero—all properties that are assumed, either explicitly or implicitly, in traditional pdf-based density ratio estimation. For each of the statements, we show how the existence of ratio limits relates to the underlying local intrinsic dimensionalities of the two functions involved, in a natural way.

3.1 Limits of Ratios of Different Functions

Intuitively speaking, the first result to be presented shows that for a density ratio limit to be greater than zero, the LID values of the numerator function and denominator function must be the same—any difference in the LID values would result in the limit either vanishing or diverging.

Lemma 1. *Let $\alpha, \beta : \mathbb{R}^{\geq 0} \rightarrow \mathbb{R}^{\geq 0}$ be functions such that $\alpha(0) = \beta(0) = 0$, and for some value of $r > 0$, their restrictions to the interval $(0, r)$ are continuously differentiable and positive. Let us also assume that ID_α^* and ID_β^* both exist and are positive. Further, let $\Delta = ID_\beta^* - ID_\alpha^*$, and let β_0 be the function obtained by decomposing β into $\beta(u) = u^\Delta \cdot \beta_0(u)$. Consider the limits of ratios*

$$\lambda \triangleq \lim_{u \rightarrow 0} \frac{\beta(u)}{\alpha(u)} \quad \text{and} \quad \lambda_0 \triangleq \lim_{u \rightarrow 0} \frac{\beta_0(u)}{\alpha(u)}.$$

If λ exists and is positive, then $ID_\alpha^ = ID_\beta^*$. Alternatively, if λ_0 exists and is positive, then the following statements hold:*

1. $\lambda = 0$ if and only if $ID_\alpha^* < ID_\beta^*$ (in which case $\Delta > 0$);
2. $\lambda > 0$ if and only if $ID_\alpha^* = ID_\beta^*$ (in which case $\Delta = 0$ and $\lambda = \lambda_0$);
3. λ diverges to $+\infty$ if and only if $ID_\alpha^* > ID_\beta^*$ (in which case $\Delta < 0$).

Proof. For the limit λ_0 to exist, $\alpha(0) = 0$ implies that $\beta_0(0) = 0$ as well. Note also that β_0 must be continuously differentiable and positive over the full range $(0, r)$. L'Hôpital's rule can therefore be applied together with Theorem 1, yielding

$$\lambda_0 = \lim_{u \rightarrow 0} \frac{\beta_0(u)}{\alpha(u)} = \lim_{u \rightarrow 0} \frac{\beta'_0(u)}{\alpha'(u)} = \lim_{u \rightarrow 0} \frac{\text{ID}_{\beta_0}(u) \cdot \beta_0(u)}{\text{ID}_{\alpha}(u) \cdot \alpha(u)}.$$

The right-hand limit can be separated to produce

$$\lambda_0 = \lim_{u \rightarrow 0} \frac{\beta_0(u)}{\alpha(u)} = \frac{\text{ID}_{\beta_0}^*}{\text{ID}_{\alpha}^*} \lim_{u \rightarrow 0} \frac{\beta_0(u)}{\alpha(u)} = \frac{\text{ID}_{\beta_0}^*}{\text{ID}_{\alpha}^*} \cdot \lambda_0.$$

These equalities imply that whenever $\lambda_0 > 0$, we have that $\text{ID}_{\alpha}^* = \text{ID}_{\beta_0}^*$, and also that $\text{ID}_{\beta}^* = \text{ID}_{\beta_0}^* + \Delta$. Similar arguments also show that $\text{ID}_{\alpha}^* = \text{ID}_{\beta}^*$ whenever λ exists and is positive.

The existence of limit λ_0 has implications for the existence of λ , which can be expressed as

$$\lambda = \lim_{u \rightarrow 0} \frac{\beta(u)}{\alpha(u)} = \lim_{u \rightarrow 0} \frac{u^{\Delta} \cdot \beta_0(u)}{\alpha(u)} = \lambda_0 \lim_{u \rightarrow 0} u^{\Delta}.$$

Whenever λ_0 exists and is positive, we have that λ behaves as the limit of u^{Δ} as $u \rightarrow 0$, in that $\lambda = 0$ if and only if $\Delta > 0$, and λ diverges to $+\infty$ if and only if $\Delta < 0$. Otherwise, $\lambda > 0$ if and only if $\Delta = 0$, in which case $\lambda = \lambda_0$. \square

Lemma 1 tells us that in an asymptotic sense, density ratios are only meaningful when the local intrinsic dimensionalities of the two functions are equal. This accords well with what is already known of standard density ratios as defined using probability densities. However, the LID model has a distinct advantage in that it is implicitly defined for any locality that admits a ‘local’ distance distribution, and that when modeling a ‘global’ distribution, no arbitrary universally-fixed local intrinsic dimensionality need be imposed.

3.2 Parameterized Limits Involving a Single Function

Next, we turn our attention to the effect of applying a common function to the numerator and denominator of an existing ratio. Equivalently, this situation can be regarded as the effect on an existing ratio of two parameterized values of the same function, as the common parameter tends to zero. Here, we state and prove results showing that the LID Representation Theorem applies with simplified conditions, in terms of the existence of a limit involving the parameterization itself.

Lemma 2. *Let $F : \mathbb{R}^{\geq 0} \rightarrow [0, 1]$ be a non-decreasing function, and assume that ID_F^* exists. Let $\alpha, \beta : \mathbb{R}^{\geq 0} \rightarrow \mathbb{R}^{\geq 0}$ be functions such that $\alpha(0) = \beta(0) = 0$, and for some value of $r > 0$, their restrictions to the interval $[0, r)$ are continuously differentiable and strictly monotonically increasing. For any constant $c \neq 0$,*

$$\lim_{u \rightarrow 0} \left(\frac{\beta(u)}{\alpha(u)} \right)^c \cdot G_F(\alpha(u), \beta(u)) = \lim_{u \rightarrow 0} \left(\frac{\beta(u)}{\alpha(u)} \right)^c = \lambda^c \tag{1}$$

whenever the limit $\lambda = \lim_{u \rightarrow 0} \frac{\beta(u)}{\alpha(u)}$ exists and is positive. If instead $c > 0$ and λ diverges to $+\infty$, or if $c < 0$ and $\lambda = 0$, then the limits in Eq. 1 both diverge to $+\infty$. Otherwise, if $c > 0$ and $\lambda = 0$, or if $c < 0$ and λ diverges to $+\infty$, then the limits in Eq. 1 are both zero.

Proof. Since ID_F^* is assumed to exist, for any real value $\epsilon \in (0, r)$ there must exist a value $0 < \delta < \epsilon < r$ such that $t < \delta$ implies that $|\text{ID}_F(t) - \text{ID}_F^*| < \epsilon$. Therefore, when $0 < \alpha(u) < \delta$ and $0 < \beta(u) < \delta$,

$$\begin{aligned} \ln G_F(\alpha(u), \beta(u)) &= \int_{\alpha(u)}^{\beta(u)} \frac{\text{ID}_F^* - \text{ID}_F(t)}{t} dt \\ |\ln G_F(\alpha(u), \beta(u))| &\leq \epsilon \cdot \left| \int_{\alpha(u)}^{\beta(u)} \frac{1}{t} dt \right| = \epsilon \cdot \left| \ln \frac{\beta(u)}{\alpha(u)} \right|. \end{aligned}$$

Exponentiating, and multiplying through by $(\beta(u)/\alpha(u))^c$, we obtain

$$\left(\frac{\beta(u)}{\alpha(u)}\right)^{c-\epsilon_0} \leq \left(\frac{\beta(u)}{\alpha(u)}\right)^c \cdot G_F(\alpha(u), \beta(u)) \leq \left(\frac{\beta(u)}{\alpha(u)}\right)^{c+\epsilon_0},$$

where $\epsilon_0 = \epsilon$ if $\beta(u) \geq \alpha(u)$, and $\epsilon_0 = -\epsilon$ otherwise.

Let us assume that the limit $\lambda = \lim_{u \rightarrow 0} \beta(u)/\alpha(u)$ exists. If $\lambda > 0$, then $\lim_{u \rightarrow 0} (\beta(u)/\alpha(u))^c$ also exists, and equals λ^c . In this case, as ϵ and δ tend to 0, the monotonicity of $\alpha(u)$ and $\beta(u)$ implies that u is driven to 0 as well. Thus, we have that

$$\lim_{u \rightarrow 0} \left(\frac{\beta(u)}{\alpha(u)}\right)^c \cdot G_F(\alpha(u), \beta(u)) = \lambda^c = \lim_{u \rightarrow 0} \left(\frac{\beta(u)}{\alpha(u)}\right)^c.$$

However, if $\lambda = 0$ or diverges to $+\infty$, then similar arguments show that the limits in Eq. 1 both diverge to $+\infty$ (if $c > 0$ and λ diverges, or $c < 0$ and $\lambda = 0$) or both converge to 0 (if $c > 0$ and $\lambda = 0$, or $c < 0$ and λ diverges). \square

Lemma 2 states conditions for which the LID representation function G_F can be ignored. Using the lemma leads to the following simplified restatement of the LID Representation Theorem itself.

Theorem 3. *Let $F : \mathbb{R}^{\geq 0} \rightarrow [0, 1]$ be a non-decreasing function, and assume that ID_F^* exists and is positive. Let $\alpha, \beta : \mathbb{R}^{\geq 0} \rightarrow \mathbb{R}^{\geq 0}$ be functions such that $\alpha(0) = \beta(0) = 0$, and for some value of $r > 0$, their restrictions to the interval $[0, r)$ are continuously differentiable and strictly monotonically increasing. Then*

$$\lim_{u \rightarrow 0} \frac{F(\beta(u))}{F(\alpha(u))} = \lim_{u \rightarrow 0} \left(\frac{\beta(u)}{\alpha(u)}\right)^{\text{ID}_F^*} = \lambda^{\text{ID}_F^*} \tag{2}$$

whenever the limit $\lambda = \lim_{u \rightarrow 0} \frac{\beta(u)}{\alpha(u)}$ exists. If instead λ diverges to $+\infty$, then the limits in Eq. 2 both diverge to $+\infty$.

Proof. Applying Theorem 2 to the first limit in Eq. 2 yields

$$\lim_{u \rightarrow 0} \frac{F(\beta(u))}{F(\alpha(u))} = \lim_{u \rightarrow 0} \left(\frac{\beta(u)}{\alpha(u)} \right)^{\text{ID}_F^*} \cdot G_F(\alpha(u), \beta(u)).$$

The result then follows from Lemma 2, with $c = \text{ID}_F^*$. □

3.3 Limits of Ratios of Inverse Functions

We next show the effect of taking the limit of ratios of the inverse of functions, in terms of the original functions as well as their local intrinsic dimensionalities. Here, we find that the limit of the ratio of the functions equals that of the ratio of their inverses, raised to the power of the local intrinsic dimension of either function.

When the original function limits are interpreted as the probability measure captured within neighborhoods whose radii tend to zero, the limits of their inverse functions can be regarded as the radii associated with neighborhoods whose captured probability measure tends to zero (as the neighborhoods shrink). Assessing the limits of ratios involving inverse functions thus gives modelers greater flexibility in designing estimators for density ratios.

Theorem 4. *Let $\alpha, \beta : \mathbb{R}^{\geq 0} \rightarrow \mathbb{R}^{\geq 0}$ be functions such that $\alpha(0) = \beta(0) = 0$, and for some value of $r > 0$, their restrictions to the interval $[0, r)$ are continuously differentiable and strictly monotonically increasing. Let us also assume that ID_α^* and ID_β^* both exist and are positive. If the limit ratio $\lambda \triangleq \lim_{u \rightarrow 0} \frac{\beta(u)}{\alpha(u)}$ exists, then*

$$\lambda = \lim_{p \rightarrow 0} \left(\frac{\alpha^{-1}(p)}{\beta^{-1}(p)} \right)^{\text{ID}_\alpha^*} = \lim_{p \rightarrow 0} \left(\frac{\alpha^{-1}(p)}{\beta^{-1}(p)} \right)^{\text{ID}_\beta^*}. \tag{3}$$

In addition, the limits in Eq. 3 diverge to $+\infty$ whenever λ diverges to $+\infty$.

Proof. Since $\alpha(0) = \beta(0) = 0$, the strict monotonicity of α and β over the range $[0, r)$ imply that their inverse functions $\alpha^{-1}(p)$ and $\beta^{-1}(p)$ both exist for all $p \in [0, s)$, where $s \triangleq \min\{\alpha(r), \beta(r)\}$. Moreover, the inverse function theorem implies that α^{-1} and β^{-1} are themselves continuously differentiable over $[0, s)$.

Consequently, using Theorem 3, the limit ratio λ can be expanded as follows:

$$\begin{aligned} \lambda &= \lim_{u \rightarrow 0} \frac{\beta(u)}{\alpha(u)} = \lim_{u \rightarrow 0} \frac{\alpha(\alpha^{-1}(\beta(u)))}{\alpha(\beta^{-1}(\beta(u)))} \\ &= \lim_{p \rightarrow 0} \left(\frac{\alpha^{-1}(p)}{\beta^{-1}(p)} \right)^{\text{ID}_\alpha^*} = \lim_{p \rightarrow 0} \left(\frac{\alpha^{-1}(p)}{\beta^{-1}(p)} \right)^{\text{ID}_\beta^*}, \end{aligned}$$

after the substitution of $p = \beta(u)$. Hence, Theorem 3 guarantees that all these limits exist and equal λ if the limit λ exists; otherwise, if λ diverges to $+\infty$, then the limits all diverge. Note that the last equality holds due to Lemma 1, which states that $\text{ID}_\alpha^* = \text{ID}_\beta^*$ for the case when $\lambda > 0$. □

It is worth noting here that, as shown by Lemma 1, these limits are only positive when the LID values of the two original functions are identical; otherwise, the limits either both vanish or both diverge.

4 Conclusion

In this paper, we presented theoretical statements that can serve as a foundation for modeling density ratios in terms of local intrinsic dimensionality. These formulations give greater flexibility when modeling data under the assumption of local variation in intrinsic dimensionality, in that no explicit dependence on a fixed-dimensional data representation is required. In particular, the result of Theorem 4 on the existence and nature of the ratio of inverse functions allows modelers to move flexibly and interchangeably between a distance-based model of local probability density on the one hand, and a probability-based model of neighborhood radius ratios on the other—all taking into proper account the effect of local intrinsic dimensionality.

As mentioned earlier, distributional modeling of density ratios is already a well-established strategy within the machine learning community [34]. The density ratio formulation considered in this paper can (in principle) be substituted for conventional pdf-based density ratios in models involving smooth (continuously differentiable) distributions—the ratio of pdfs thereby being replaced by the limits of ratios of cdfs of distance distributions. In data mining and other settings where the data is not explicitly modeled in terms of smooth distributions, the connection to LID-aware density ratios can still be made, albeit less directly. A data set consisting of n points can be regarded as a sample drawn from some unknown global distribution that, from any given point of interest \mathbf{x} , induces a distribution of distances. Letting F denote the cdf of the distribution of distances to \mathbf{x} , the expected number of data points lying within distance r of \mathbf{x} is simply $n \cdot F(r)$. If r is the radius of the k -nearest neighborhood of \mathbf{x} , then k/n can serve as a (crude) estimate of $F(r)$. In this sense, within the underlying continuous model, formulations involving k , n and r can be viewed as sample-based estimators of formulations involving F and ID_F . Continuous limit-based and LID-aware forms such as those considered in this paper can consequently be applied to yield explanations of asymptotic behavior, as the sample size n tends to infinity. Moreover, an LID-aware asymptotic analysis could conceivably suggest alternative heuristics with improved performance characteristics.

In the data mining setting of outlier detection [7], density ratio modeling has already assumed a place of central importance. The classic (and still state-of-the-art) LOF family of outlier detection methods [5] assess the outlierness of a test data point through a ratio of densities, one at the test point, and the other with respect to an aggregation of points in the vicinity of the test point. LOF formulations essentially estimate density using neighborhood-based criteria, in terms of the radii within which a predetermined number of points are captured. However, for those settings where the dimensional characteristics are assumed to vary from locality to locality, Theorem 4 implies that any use of the

ratio of distances to estimate density ratios should also take the local intrinsic dimensionality into account. The theoretical results of this paper underscore the need for the development of LID-aware local outlier detection techniques; work in this direction is already underway.

LID-aware density ratio modeling may also have useful applications in density-based clustering [6] and other non-parametric unsupervised learning settings that exploit similarity information. One such possibility involves the well-known DBSCAN family of clustering methods [12], which relies on absolute density thresholding to determine clusters, where density is (typically) estimated in terms of the numbers of points enclosed within neighborhoods of fixed radius. Extension of the DBSCAN strategy to account for LID-aware density would potentially allow for the formation of clusters whose densities are not necessarily high in an absolute sense, but instead high relative to the density within some background distribution. This has the advantage that locally-dense configurations of data points within sparse regions may still be discoverable as clusters, even in the presence of other clusters in regions that are much more dense.

In the larger sense, with its unification of the notions of distance, probability, and local dimensionality, LID-aware modeling presents an opportunity to reconcile established heuristics in the area of data mining with the distributional framework that serves as the foundation of machine learning and other areas in AI. By breaking the dependence of models on a global dimension parameter, it also offers better support in subspace-based modeling and other contexts in which sparse-featured solutions are sought. The theoretical results presented in this paper on LID-aware density ratios will hopefully encourage and support further research in this direction.

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

References

1. Amsaleg, L., et al.: Extreme-value-theoretic estimation of local intrinsic dimensionality. *Data Min. Knowl. Discov.* **32**(6), 1768–1805 (2018)
2. Amsaleg, L., Chelly, O., Houle, M.E., Kawarabayashi, K., Radovanović, M., Treeratatanajaru, W.: Intrinsic dimensionality within tight localities. In: *Proceedings of the 19th SIAM International Conference on Data Mining (SDM)*, Calgary, AB, Canada, pp. 181–189 (2019)
3. Aumüller, M., Ceccarello, M.: The role of local intrinsic dimensionality in benchmarking nearest neighbor search. In: Amato, G., Gennaro, C., Oria, V., Radovanović, M. (eds.) *SISAP 2019*. LNCS, vol. 11807, pp. 113–127. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32047-8_11
4. Becker, R., Hafnaoui, I., Houle, M.E., Li, P., Zimek, A.: Subspace determination through local intrinsic dimensional decomposition. In: Amato, G., Gennaro, C., Oria, V., Radovanović, M. (eds.) *SISAP 2019*. LNCS, vol. 11807, pp. 281–289. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32047-8_25
5. Breunig, M.M., Kriegel, H.P., Ng, R., Sander, J.: LOF: identifying density-based local outliers. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Dallas, TX, pp. 93–104 (2000)

6. Campello, R.J.G.B., Kröger, P., Sander, J., Zimek, A.: Density-based clustering. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **10**(2) (2020)
7. Campos, G.O., et al.: On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data Min. Knowl. Discov.* **30**(4), 891–927 (2016)
8. Casanova, G., et al.: Dimensional testing for reverse k -nearest neighbor search. *Proc. VLDB Endow.* **10**(7), 769–780 (2017)
9. Cochran, W.G.: *Sampling Techniques*. Wiley, Hoboken (1977)
10. Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*. Springer, London (2001). <https://doi.org/10.1007/978-1-4471-3675-0>
11. de Vries, T., Chawla, S., Houle, M.E.: Density-preserving projections for large-scale local anomaly detection. *Knowl. Inf. Syst. (KAIS)* **32**(1), 25–52 (2012)
12. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, Portland, OR, pp. 226–231 (1996)
13. Hashem, T., Rashidi, L., Bailey, J., Kulik, L.: Characteristics of local intrinsic dimensionality (LID) in subspaces: local neighbourhood analysis. In: Amato, G., Gennaro, C., Oria, V., Radovanović, M. (eds.) *SISAP 2019*. LNCS, vol. 11807, pp. 247–264. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32047-8_22
14. Hill, B.M.: A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **3**(5), 1163–1174 (1975)
15. Houle, M.E.: Dimensionality, discriminability, density and distance distributions. In: *Proceedings of the 13th IEEE International Conference on Data Mining Workshops, ICDM Workshops*, Dallas, TX, pp. 468–473 (2013)
16. Houle, M.E.: Local intrinsic dimensionality I: an extreme-value-theoretic foundation for similarity applications. In: Beecks, C., Borutta, F., Kröger, P., Seidl, T. (eds.) *SISAP 2017*. LNCS, vol. 10609, pp. 64–79. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68474-1_5
17. Houle, M.E.: Local intrinsic dimensionality II: multivariate analysis and distributional support. In: Beecks, C., Borutta, F., Kröger, P., Seidl, T. (eds.) *SISAP 2017*. LNCS, vol. 10609, pp. 80–95. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68474-1_6
18. Houle, M.E., Kashima, H., Nett, M.: Generalized expansion dimension. In: *ICDM Workshop Practical Theories for Exploratory Data Mining (PTDM)*, pp. 587–594 (2012)
19. Houle, M.E., Ma, X., Nett, M., Oria, V.: Dimensional testing for multi-step similarity search. In: *Proceedings of the 12th IEEE International Conference on Data Mining (ICDM)*, Brussels, Belgium, pp. 299–308 (2012)
20. Houle, M.E., Ma, X., Oria, V.: Effective and efficient algorithms for flexible aggregate similarity search in high dimensional spaces. *IEEE Trans. Knowl. Data Eng.* **27**(12), 3258–3273 (2015)
21. Houle, M.E., Ma, X., Oria, V., Sun, J.: Efficient algorithms for similarity search in axis-aligned subspaces. In: Traina, A.J.M., Traina, C., Cordeiro, R.L.F. (eds.) *SISAP 2014*. LNCS, vol. 8821, pp. 1–12. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11988-5_1
22. Houle, M.E., Nett, M.: Rank-based similarity search: reducing the dimensional dependence. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(1), 136–150 (2015)

23. Houle, M.E., Oria, V., Wali, A.M.: Improving k -NN graph accuracy using local intrinsic dimensionality. In: Beecks, C., Borutta, F., Kröger, P., Seidl, T. (eds.) SISAP 2017. LNCS, vol. 10609, pp. 110–124. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68474-1_8
24. Houle, M.E., Schubert, E., Zimek, A.: On the correlation between local intrinsic dimensionality and outlieriness. In: Marchand-Maillet, S., Silva, Y.N., Chávez, E. (eds.) SISAP 2018. LNCS, vol. 11223, pp. 177–191. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02224-2_14
25. Karamata, J.: Sur un mode de croissance régulière, théorèmes fondamentaux. Bull. Soc. Math. France **61**, 55–62 (1933)
26. Karger, D.R., Ruhl, M.: Finding nearest neighbors in growth-restricted metrics. In: Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC), Montreal, QC, Canada, pp. 741–750 (2002)
27. Kullback, S.: Information Theory and Statistics. Wiley, Hoboken (1959)
28. Liese, F., Vajda, I.: On divergences and informations in statistics and information theory. IEEE Trans. Inf. Theory **52**(10), 4394–4412 (2006)
29. Ma, X., et al.: Characterizing adversarial subspaces using local intrinsic dimensionality, pp. 1–15 (2018)
30. Ma, X., et al.: Dimensionality-driven learning with noisy labels, pp. 3361–3370 (2018)
31. Romano, S., Chelly, O., Nguyen, V., Bailey, J., Houle, M.E.: Measuring dependency via intrinsic dimensionality. In: ICPR16, pp. 1207–1212, December 2016
32. Schubert, E., Gertz, M.: Intrinsic t -stochastic neighbor embedding for visualization and outlier detection. In: Beecks, C., Borutta, F., Kröger, P., Seidl, T. (eds.) SISAP 2017. LNCS, vol. 10609, pp. 188–203. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68474-1_13
33. Scott, A.J., Wu, C.: On the asymptotic distribution of ratio and regression estimators. J. Am. Stat. Assoc. **76**, 98–102 (2006)
34. Sugiyama, M., Suzuki, T., Kanamori, T.: Density Ratio Estimation in Machine Learning. Cambridge University Press, Cambridge (2012)
35. Terrell, D.G., Scott, D.W.: Variable kernel density estimation. Ann. Stat. **20**(3), 1236–1265 (1992)