# Measuring Dependency via Intrinsic Dimensionality

## Supplementary Material

### Proofs of Propositions and Theorems

**Theorem 1.** *Let $X$ be a set of $D$ continuous variables, $f(x)$ the p.d.f. of the distribution from which $X$ is drawn, and $\mathrm{ID}(x)$ the local intrinsic dimension at the locality $x$. The $\alpha$-Rényi dimension can be expressed as:*

$$\dim_\alpha(X) = \frac{\int f^\alpha(x)\,\mathrm{ID}(x)\,\mathrm{d}x}{\int f^\alpha(x)\,\mathrm{d}x}.$$

*Proof.* We first note that the following holds true for the generalized correlation integral in Equation (4):

$$C_\alpha(X,r) = \left(\int \left(\int f(y)\bar{\mathbf{1}}(x,y,r)\,\mathrm{d}y\right)^{\alpha-1} f(x)\,\mathrm{d}x\right)^{\frac{1}{\alpha-1}} = \left(\int F_R^{\alpha-1}(x,r) f(x)\,\mathrm{d}x\right)^{\frac{1}{\alpha-1}},$$

where $F_R(x,r) = \int f(y)\bar{\mathbf{1}}(x,y,r)\,\mathrm{d}y$ is the number of points at distance smaller than $r$ from $x$. Then, we use l'Hôpital's rule on the definition of $\dim_\alpha(X)$ in Equation (3):

$$\begin{aligned}
\dim_\alpha(X) &= \lim_{r\to 0^+} \frac{\log\left(\int F_R^{\alpha-1}(x,r)f(x)\,\mathrm{d}x\right)}{(\alpha-1)\log r}\\
&\overset{\text{H.}}{=} \lim_{r\to 0^+} \frac{r\int(\alpha-1)F_R^{\alpha-2}(x,r)f_R(x,r)f(x)\,\mathrm{d}x}{(\alpha-1)\int F_R^{\alpha-1}(x,r)f(x)\,\mathrm{d}x}\\
&= \lim_{r\to 0^+} \frac{\int F_R^{\alpha-1}(x,r)\frac{rf_R(x,r)}{F_R(x,r)}f(x)\,\mathrm{d}x}{\int F_R^{\alpha-1}(x,r)f(x)\,\mathrm{d}x}\\
&= \lim_{r\to 0^+} \frac{\int F_R^{\alpha-1}(x,r)\mathrm{ID}(x,r)f(x)\,\mathrm{d}x}{\int F_R^{\alpha-1}(x,r)f(x)\,\mathrm{d}x}.
\end{aligned}$$

As $r$ tends to $0^+$, $F_R(x,r)$ tends to $f(x)$. Therefore:

$$\dim_\alpha(X) = \frac{\int f^\alpha(x)\mathrm{ID}(x)\,\mathrm{d}x}{\int f^\alpha(x)\,\mathrm{d}x}$$

□

**Theorem 2.** *The kNN estimator of $\dim_\alpha(X)$ is:*

$$\widehat{\dim_\alpha}(X) = \frac{\sum_{i=1}^n \widehat{\mathrm{ID}}(x_i)(d_k(x_i)^{-D})^{\alpha-1}}{\sum_{i=1}^n (d_k(x_i)^{-D})^{\alpha-1}}.$$

*Proof.* We first prove a more general result: if $K(\cdot)$ is a kernel function with width $h$, then for $\alpha \geq 1$,

$$\widehat{\dim_\alpha}(X) = \frac{\sum_{i=1}^n \widehat{\mathrm{ID}}(x_i)\Big(\sum_{j=1}^n K(\|x_i - x_j\|, h)\Big)^{\alpha-1}}{\sum_{i=1}^n \Big(\sum_{j=1}^n K(\|x_i - x_j\|, h)\Big)^{\alpha-1}}. \tag{8}$$

To prove this, note that for $\alpha \geq 1$, $\dim_\alpha(X) = \frac{\int f(x)f(x)^{\alpha-1}\mathrm{ID}(x)\,\mathrm{d}x}{\int f(x)f(x)^{\alpha-1}\,\mathrm{d}x}$. The p.d.f. $f(x)$ of $X$ can be estimated with kernel functions $K(\cdot)$ via summation over all data points $x_i$: $\hat{f}(x) = \frac{1}{n}\sum_{j=1}^n \frac{1}{h}K(\|x - x_j\|, h)$. If we have a reliable sample of $n$ i.i.d data points from $X$, the expected value $\int f(x)g(x)\,\mathrm{d}x$ of any function $g(x)$ over the p.d.f. $f(x)$ can be estimated using the formula: $\frac{1}{n}\sum_{i=1}^n g(x_i)$. Therefore the denominator of $\dim_\alpha(X)$ can be estimated with $\frac{1}{n}\sum_{i=1}^n \hat{f}(x_i)^{\alpha-1} = \frac{1}{n}\sum_{i=1}^n(\frac{1}{n}\sum_{j=1}^n \frac{1}{h}K(\|x_i - x_j\|, h))^{\alpha-1}$. The numerator is instead equal to $\frac{1}{n}\sum_{i=1}^n \widehat{\mathrm{ID}}(x_i)(\frac{1}{n}\sum_{j=1}^n \frac{1}{h}K(\|x_i - x_j\|, h))^{\alpha-1}$. The formula in Eq. (8) can be easily obtained with algebraic simplifications.

With regards to the $k$NN estimator, it is possible to prove that $K(\|x_i - x_j\|) = \frac{\mathbf{1}(\|x_i - x_j\| \leq r)}{V_D(r)}$ is a proper kernel, where $r$ is a given radius and $V_D(r) = \frac{\pi^{D/2}}{\Gamma(D/2+1)}r^D$ is the volume of a $D$-dimensional sphere with radius $r$. A valid choice for the radius $r$ is the distance $d_k(x_i)$ from $x_i$ to its $k$th nearest neighbor. Given that the number of data points at distance less than or equal to $d_k(x_i)$ from $x_i$ is exactly $k$, we have $\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{1}(\|x_i - x_j\| \leq d_k(x_j))}{V_D(d_k(x_j))} = \frac{1}{n}\frac{k}{V_D(d_k(x_j))} = \frac{1}{n}\frac{k\Gamma(D/2+1)d_k(x_i)^{-D}}{\pi^{D/2}}$. The result follows from algebraic manipulations. □

**Proposition 1.** *Let $X$ be a set of $D$ continuous variables:*

1) $0 \leq \mathrm{IDD}(X) \leq 1$*;*
2) $\mathrm{IDD}(X) = 0$ *iff all $X_i$ are independent;*
3) $\mathrm{IDD}(X) = 1$ *if there exist one or more manifolds of dimension 1 whose union embeds $X$;*
4) $\mathrm{IDD}(X) = 1$ *if there exists $1 \leq i \leq D$ such that for all $j \neq i$, $X_j$ is a a function or multivalued function of $X_i$.*

*Proof.*
Point 1: By definition, $\dim(X) = \lim_{\delta\to 0^+} \frac{H(X,\delta)}{\log 1/\delta}$. Then regarding the lower bound of IDD, $\sum_{i=1}^D \dim(X_i) - \dim(X)$ is equal to:

$$\begin{aligned}
&= \sum_{i=1}^D \lim_{\delta\to 0^+} \frac{H(X_i,\delta)}{\log 1/\delta} - \lim_{\delta\to 0^+} \frac{H(X,\delta)}{\log 1/\delta}\\
&= \lim_{\delta\to 0^+} \frac{1}{\log 1/\delta}\Big(\sum_{i=1}^D H(X_i,\delta) - H(X,\delta)\Big)\\
&= \lim_{\delta\to 0^+} \frac{1}{\log 1/\delta}\mathrm{KL}\Big(p_X(x,\delta)\|p_{X_1}(x_1,\delta)\cdots p_{X_D}(x_D,\delta)\Big),
\end{aligned}$$

where KL is the Kullback-Leibler divergence, which is greater or equal to 0 for any $\delta > 0$. Regarding the upper bound of IDD, we use the known fact that the Shannon entropy satisfies $H(X) \geq \max_i H(X_i)$ to prove the following inequalities for

$\sum_{i=1}^{D} \dim(X_i) - \dim(X)$:

$$
\begin{aligned}
&= \sum_{i=1}^{D} \lim_{\delta\to0^+} \frac{H(X_i,\delta)}{\log 1/\delta} - \lim_{\delta\to0^+} \frac{H(X,\delta)}{\log 1/\delta} \\
&\le \sum_{i=1}^{D} \lim_{\delta\to0^+} \frac{H(X_i,\delta)}{\log 1/\delta} - \lim_{\delta\to0^+} \frac{\max_i H(X_i,\delta)}{\log 1/\delta} \\
&= \sum_{i=1}^{D} \lim_{\delta\to0^+} \frac{H(X_i,\delta)}{\log 1/\delta} - \max_i \lim_{\delta\to0^+} \frac{H(X_i,\delta)}{\log 1/\delta} \\
&= \sum_{i=1}^{D} \lim_{\delta\to0^+} \frac{H(X_i,\delta)}{\log 1/\delta} - \max_i \dim(X_i).
\end{aligned}
$$

Since the Shannon entropy is a continuous function, and since $X$ is continuous, it is possible to interchange the limit and max operations.
Point 2: As shown for Point 1 above, $\sum_{i=1}^{D} \dim(X_i) - \mathrm{IDD}(X)$ is equal to

$$
\lim_{\delta\to0^+} \frac{1}{\log 1/\delta} \mathrm{KL}\Big(p_X(x,\delta) \| p_{X_1}(x_1,\delta)\cdots p_{X_D}(x_D,\delta)\Big).
$$

The result follows from the fact that for any $\delta > 0$, the KL divergence is equal to 0 iff all variables $X$ are independent.
Point 3: If there exist at least a manifold of dimension 1 embedded in $X$, then $\mathrm{ID}(x) = 1$ for any locality $x$. With $\dim(X) = 1$ being the expected ID over the p.d.f. of $X$, we have that $\mathrm{DID}(X) = 1$. According to Theorem 1 in [10] if $X_i$ is a continuous random variable, $\dim(X_i) = 1$. Given that we are considering continuous random variables $X_i$, $\max_i \dim(X_i) = 1$, and therefore $\mathrm{IDD}(X) = 1$.
Point 4: follows immediately from Point 3. □