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

\[ C_\alpha(X, r) = \left( \int \left( \int f(y) \mathbb{1}(x, y, r) \, dy \right)^{\alpha - 1} f(x) \, dx \right)^{-1} \]

\[ = \left( \int F_R^{\alpha - 1}(x, r) f(x) \, dx \right)^{-1}, \]

where \( F_R(x, r) = \int f(y) \mathbb{1}(x, y, r) \, dy \) 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 (5):

\[ \dim_{\alpha}(X) = \lim_{r \to 0^+} \log \left( \frac{\int F_R^{\alpha - 1}(x, r) f(x) \, dx}{\alpha - 1} \right) \]

\[ = \lim_{r \to 0^+} \frac{r \int (\alpha - 1) F_R^{\alpha - 2}(x, r) f(x) \, dx}{\int F_R^{\alpha - 1}(x, r) f(x) \, dx} \]

\[ = \lim_{r \to 0^+} \frac{\int F_R^{\alpha - 1}(x, r) f(x) \, dx}{\int F_R^{\alpha - 1}(x, r) \mathbb{1}(x, r) \, dx} \]

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

\[ \dim_{\alpha}(X) = \frac{\int f(x) \mathbb{1}(x, r) \, dx}{\int f(x) \, dx}. \]

Theorem 2. The kNN estimator of \( \dim_{\alpha}(X) \) is:

\[ \widehat{\dim}_{\alpha}(X) = \frac{\sum_{i=1}^n \widehat{\text{ID}}(x_i)(d_k(x_i)^{-(D-\alpha)})^{-1}}{\sum_{i=1}^n (d_k(x_i)^{-(D-\alpha)})^{-1}}. \]

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

\[ \widehat{\dim}_{\alpha}(X) = \frac{\sum_{i=1}^n \widehat{\text{ID}}(x_i) \left( \sum_{j=1}^n K(\|x_i - x_j\|, h) \right)^{-1}}{\sum_{i=1}^n \left( \sum_{j=1}^n K(\|x_i - x_j\|, h) \right)^{-1}}. \]

To prove this, note that for \( \alpha \geq 1 \), \( \dim_{\alpha}(X) = \frac{\int f(x) f(x)^{\alpha - 1} \mathbb{1}(x, r) \, dx}{\int f(x) f(x)^{\alpha - 1} \, dx} \). The p.d.f. \( f(x) \) of \( X \) can be estimated with kernel functions \( K(\cdot) \) via summation over all data points \( x_i \):

\[ f(x) = \frac{1}{n} \sum_{j=1}^n \frac{1}{\pi d_k(x_j)^2} K(\|x - x_j\|, h). \]

If we have a reliable sample of \( n \) i.i.d. data points from \( X \), the expected value \( \int f(x) g(x) \, dx \) 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 \( \frac{1}{n} \sum_{i=1}^n f(x_i)^{\alpha - 1} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{n} \sum_{j=1}^n K(\|x_i - x_j\|, h) \right)^{\alpha - 1} \). The numerator is instead equal to \( \frac{1}{n} \sum_{i=1}^n \widehat{\text{ID}}(x_i) \left( \frac{1}{n} \sum_{j=1}^n \frac{1}{\pi d_k(x_j)^2} K(\|x_i - x_j\|, h)^{\alpha - 1} \right) \). The formula in Eq. (9) can be easily obtained with algebraic simplifications.

With regards to the kNN estimator, it is possible to prove that \( K(\|x_i - x_j\|) = \frac{1}{\pi d_k(x_j)^2} \mathbb{1}(x, r) \) is a proper kernel, where \( r \) is a given radius and \( V_D(r) = \frac{\pi^D}{D/2 + 1} \) 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 \( \frac{1}{n} \sum_{i=1}^n \frac{1}{\pi d_k(x_i)^2} \mathbb{1}(x, r) = \frac{1}{n} V_D(d_k(x_i)) = \frac{1}{n} V_D(d_k(x_i)) \). The result follows from algebraic manipulations.

Proposition 1. Let \( X \) be a set of \( D \) continuous variables:

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

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

\[ = \sum_{i=1}^D \lim_{\delta \to 0^+} \frac{H(X_i, \delta)}{\log 1/\delta} - \lim_{\delta \to 0^+} \frac{H(X, \delta)}{\log 1/\delta} \]

\[ = \lim_{\delta \to 0^+} \frac{1}{\log 1/\delta} \left( \sum_{i=1}^D H(X_i, \delta) - H(X, \delta) \right) \]

\[ = \lim_{\delta \to 0^+} \frac{1}{\log 1/\delta} \left( \sum_{i=1}^D H(X_i, \delta) - H(X, \delta) \right) \]

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
\[ \sum_{i=1}^D \text{dim}(X_i) - \dim(X) : \]
\[ = \sum_{i=1}^D \lim_{\delta \to 0^+} \frac{H(X_i, \delta)}{\log 1/\delta} - \lim_{\delta \to 0^+} \frac{H(X, \delta)}{\log 1/\delta} \]
\[ \leq \sum_{i=1}^D \lim_{\delta \to 0^+} \frac{H(X_i, \delta)}{\log 1/\delta} - \max_i \lim_{\delta \to 0^+} \frac{H(X_i, \delta)}{\log 1/\delta} \]
\[ = \sum_{i=1}^D \lim_{\delta \to 0^+} \frac{H(X_i, \delta)}{\log 1/\delta} - \max \dim(X_i). \]

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, \( \sum_{i=1}^D \text{dim}(X_i) - \text{IDD}(X) \) is equal to
\[ \lim_{\delta \to 0^+} \frac{1}{\log 1/\delta} \text{KL} \left( p_X(x, \delta) \| p_{X_1}(x_1, \delta) \cdots p_{X_D}(x_D, \delta) \right). \]

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 \( \text{ID}(x) = 1 \) for any locality \( x \). With \( \dim(X) = 1 \) being the expected ID over the p.d.f. of \( X \), we have that \( \text{IDD}(X) = 1 \). 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 \( \text{IDD}(X) = 1 \).

Point 4: follows immediately from Point 3.