Estimation of Rényi Entropy and Mutual Information Based on Generalized Nearest-Neighbor Graphs

Dávid Pál†, Barnabás Póczos‡, Csaba Szepesvári†

† Department of Computer Science, University of Alberta, Edmonton, Canada
‡ School of Computer Science, Carnegie Mellon University, Pittsburgh, USA

dpal@cs.ualberta.ca, bapoczos@cs.cmu.edu, szepesva@cs.ualberta.ca

Introduction

- We are the first to correctly prove almost sure consistency and rates of convergence (Leonenko et al., 2008a, Kozachenko and Leonenko, 1987, Wang et al., 2009).
- Computationally more efficient than minimum spanning tree estimator (Hero and Michel, 1999, Póczos et al., 2010).

Definitions

The Rényi entropy and the Rényi mutual information of order α of a real-valued random variables \( X = (X_1, X_2, \ldots, X_d) \) with joint density \( f: \mathbb{R}^d \rightarrow \mathbb{R} \) and marginal densities \( f_i: \mathbb{R} \rightarrow \mathbb{R}, 1 \leq i \leq d \) are for \( \alpha \neq 1 \) respectively defined by

\[
H_{\alpha}(f) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} f^{\alpha}(x_1, x_2, \ldots, x_d) \, dx_1 \, dx_2 \, \cdots \, dx_d,
\]

\[
I_{\alpha}(f) = \frac{1}{\alpha-1} \log \int_{\mathbb{R}^d} f^\alpha(x_1, x_2, \ldots, x_d) \left( \prod_{i=1}^d f_i(x_i) \right)^{1-\alpha} \, dx_1 \, dx_2 \, \cdots \, dx_d.
\]

The limits \( H_1(f) = \lim_{\alpha \to 1} H_{\alpha}(f) \) and \( I_1(f) = \lim_{\alpha \to 1} I_{\alpha}(f) \) are the Shannon entropy and the Shannon mutual information respectively.

Generalized Nearest Neighbor Graphs

Fix a finite non-empty set \( S \) of positive integers; e.g. \( S = \{1, 2, \ldots, k\} \) or \( S = \{k\} \). Given a finite set \( V \) of points in \( \mathbb{R}^d \) we define a generalized nearest neighbor graph \( \mathbf{NN}(V) \) as a directed graph on \( V \) where for each point \( x \in V \) and each \( i \in S \) there is an edge from \( x \) to its \( i \)-th nearest neighbor in \( V \). We define

\[
L_p(V) = \sum_{(x,y) \in \mathbf{NN}(V)} \|x - y\|^p.
\]

Theorem. (Constant \( \gamma \)) If \( \mathbf{NN} \) is an i.i.d. sample of size \( n \) from the uniform distribution over \([0,1]^d\), then for any \( p \in [0, d] \) and any \( S \) there exists a constant \( \gamma \) such that

\[
\lim_{n \to \infty} \frac{\mathbf{NN}(V)}{n} = \gamma \quad \text{a.s.}
\]

Using this theorem we can estimate the constant \( \gamma \) to arbitrary precision.

Copulas and Estimator of Mutual Information

We estimate the Rényi mutual information \( I_{\alpha}(f) \) by

\[
\hat{I}_{\alpha}(f) = \hat{H}_{\alpha}(\text{Empirical Copula}(X_{1:n})).
\]

Theorem. (Consistency and Rate for \( \hat{I}_{\alpha} \)) Let \( d \geq 3 \) and \( \alpha = 1 - p/d \in (1/2, 1) \). Let \( \mu \) be an absolutely continuous distribution over \([0,1]^d\) with density \( f \). If \( X_{1:n} = (X_1, X_2, \ldots, X_n) \) is an i.i.d. sample from \( \mu \) then

\[
\lim_{n \to \infty} \hat{I}_{\alpha}(f) = I_{\alpha}(f) \quad \text{a.s.}
\]

Moreover, if the density of the copula of \( \mu \) is Lipschitz, then for any \( \delta > 0 \) with probability at least \( 1 - \delta \),

\[
\left| \hat{I}_{\alpha}(f) - I_{\alpha}(f) \right| \leq O\left( \frac{\max\{n^{-1}, n^{-1/d-\delta} \}}{\log(n^2)} \right) \quad \text{if } 0 < p < d - 1;
\]

\[
\left| \hat{I}_{\alpha}(f) - I_{\alpha}(f) \right| \leq O\left( \frac{\max\{n^{-1}, n^{-1/d-\delta} \}}{\log(n^2)} \right) \quad \text{if } d - 1 < p < d.
\]

Copula Transform

Figure 1: Nearest neighbor graph \( \mathbf{NN}(U) \) on a sample consisting of \( n = 200 \) points drawn i.i.d. from the uniform distribution over \([0,1]^d\) and with \( S = \{1, 2, 3\} \).

Figure 2: Sample from the uniform distribution over a triangle with vertices \((0,0), (1,0), (0,1)\) and its empirical copula.

References