Mutual information is copula entropy

Jian Ma and Zengqi Sun*
Department of Computer Science, Tsinghua University, Beijing 100084, China

(Dated: August 6, 2008)

We prove that mutual information is actually negative copula entropy, based on which a method for mutual information estimation is proposed.

PACS numbers: 05.90.+m, 02.50.-r, 87.10.-e

I. INTRODUCTION

In information theory, mutual information (MI) is a difference concept with entropy.[1] In this paper, we prove with copula [2] that they are essentially same – mutual information is also a kind of entropy, called *copula entropy*. Based on this insightful result, We propose a simple method for estimating mutual information.

Copula is a theory on dependence and measurement of association.[2] Sklar [3] proved that joint distribution D can be represented with copula C and margins F in the following form:

$$D(\mathbf{x}) = C(F_1(x_1), \ldots, F_N(x_N)).$$

Derived by separating the margins from joint distribution, copula has all the dependence information of random variables, which is believed that mutual information does as well.

Here gives notation. C, c denote copula function and copula density; D, F denotes joint distribution and marginal distribution; H, I, H_c denote entropy, mutual information, and copula entropy respectively. Finally, bold letters represent vectors while normal letters single variable.

II. THEOREM AND PROOF

For clarity, we give directly the main results. Please refer to [2] and references therein for more about copula.

Definition 1 (Copula entropy)**.** *Let* $\mathbf{X}$ *be random variables with marginal function* $\mathbf{u}$ *and copula density* $c(\mathbf{u})$*. Copula entropy of* $\mathbf{X}$ *is*

$$H_c(\mathbf{x}) = -\int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u}. \tag{1}$$

Theorem 1. *Mutual Information of random variables equals to the negative entropy of their corresponding copula function:*

$$I(\mathbf{x}) = --H_c(\mathbf{x}). \tag{2}$$

Proof.

$$\begin{aligned}
I(\mathbf{x}) &= \int_{\mathbf{x}} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\prod_i p_i(x_i)} d\mathbf{x} \\
&= \int_{\mathbf{x}} c(\mathbf{u}_x) \prod_i p_i(x_i) \log c(\mathbf{u}_x) d\mathbf{x} \\
&= \int_{\mathbf{u}_x} c(\mathbf{u}_x) \log c(\mathbf{u}_x) d\mathbf{u}_x \\
&= -H_c(\mathbf{x})
\end{aligned}$$

□

Entropy is the information contained in joint density function and marginal densities, while copula entropy is the information contained in copula density. The following corollary show the relation between them.

Corollary 1.

$$H(\mathbf{x}) = \sum_i H(x_i) + H_c(\mathbf{x}) \tag{3}$$

Proof. The corollary is an instant result from the definition of mutual information and theorem 1. □

The worthy-a-thousand-words results cast insight into the inner relation between mutual information and copula, and hence builds a connection between information theory and copula theory.

III. ESTIMATING MUTUAL INFORMATION VIA COPULA

With theorem 1, we propose the methods of estimating mutual information from data. The estimation composes of two steps:

1. estimating empirical copula density;

2. estimating copula entropy.

Given samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ i.i.d. generated from $X = [x_1, \ldots, x_N]$, we can easily derive empirical copula density using empirical functions

$$F_i(x_i) = \frac{1}{T} \sum_{i=1}^{T} \chi(X_t^i \leq x_t^i) \tag{4}$$

---

*Electronic address: majian03@mails.tsinghua.edu.cn

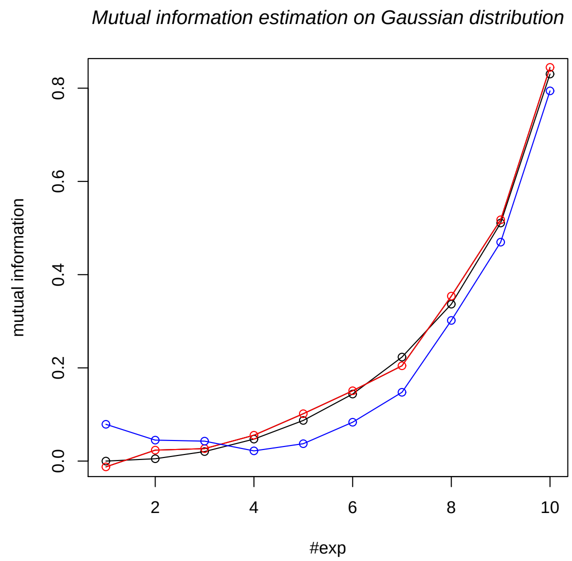

FIG. 1: Mutual information estimation. Black represents analytic value, red represents by knn, and blue represents by copula entropy.

where $i = 1, \ldots, N$, and χ is indicator function. Let $\mathbf{u} = [F_1, \ldots, F_N]$, and then we can derive a group of samples $\{\mathbf{u}_1, \ldots, \mathbf{u}_T\}$ from empirical copula density $\hat{c}(\mathbf{u})$.

Since entropy estimation is a much contributed topic, copula entropy can be achieved by well-established methods. In the following experiment, we adopt the method proposed by [4].

IV. EXPERIMENTS

To evaluate the new estimation method, we consider tow correlated standard Gaussian variables with covariance ρ, of which mutual information is $-\frac{1}{2}\log(1-\rho^2)$. In the experiments ρ is set from 0 to 0.9 with step 0.1. With each value, we generate a 1000 samples set. Copula entropy is estimated by the method in [4]. As a contrast, mutual information estimation in [4] was also run on the same sample set. Figure IV illustrates the experimental results. It can be learned that our method provides a competitive way of MI estimation.

[1] T. Cover and J. Thomas, *Elements of information theory* (Wiley New York, 1991).
[2] R. B. Nelsen, *An Introduction to Copulas*, Lecture Notes in Statistics (New York: Springer, 1999).
[3] A. Sklar, Publications de l'Institut de Statistique de l'Université de Paris **8**, 229 (1959).
[4] A. Kraskov, H. Stögbauer, and P. Grassberger, Physical Review E **69**, 066138 (2004).