

Mutual Information Is Copula Entropy

MA Jian (马健), SUN Zengqi (孙增圻)**

State Key Laboratory on Intelligent Technology and Systems,
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Abstract: Mutual information (MI) is a basic concept in information theory. Therefore, estimates of the MI are fundamentally important in most information theory applications. This paper provides a new way of understanding and estimating the MI using the copula function. First, the entropy of the copula, named the copula entropy, is defined as a measure of the dependence uncertainty represented by the copula function and then the MI is shown to be equivalent to the negative copula entropy. With this equivalence, the MI can be estimated by first estimating the empirical copula and then estimating the entropy of the empirical copula. Thus, the MI estimate is an estimation of the entropy, which reduces the complexity and computational requirements. Tests show that the method is more effective than the traditional method.

Key words: copula entropy; mutual information; estimation; empirical copula

Introduction

Since the seminal paper by Shannon^[1], information theory has been widely used in many branches of science. In information theory, entropy and mutual information (MI) are two different fundamental concepts^[2]. Entropy is defined as the measure of uncertainty associated with random variables and with multivariate MI as a measure of the mutual dependence of random variables defined as follows:

$$I(\mathbf{x}) = \sum_{i=1}^N H(x_i) - H(\mathbf{x}) \quad (1)$$

where $\mathbf{x} = [x_1, \dots, x_N] \in \mathbf{R}^N$ are random variables, H is the entropy, and I is the MI. Unlike traditional statistical measures of dependence, such as correlation (i.e. second-order statistics) and higher-order statistics^[3], MI is a measure of all orders of dependence.

Another branch of study, copula theory^[4,5], concerns the representation of dependence between random variables. Theoretically, copula is a function that

represents the dependence between random variables^[4,5]. Such a representation is guaranteed by the Sklar theorem^[6]. Let P be a joint distribution function with margins $\{F_i, i = 1, \dots, N\}$. The Sklar theorem states that P can be represented as copula function C with F as its arguments:

$$P(\mathbf{x}) = C(F_1(x_1), \dots, F_N(x_N)) \quad (2)$$

If the random variables F are separated with margins from the joint distribution function, the copula has all the dependence information of the random variables.

This paper investigates the relation between the measure of dependence, MI, and the representation of the dependence, the copula function. De la Peña et al.^[7] studied the properties of MI and other measures of dependence with the copula represented as U-statistics. Davy and Doucet^[8] studied the relation between copula and Cohen-Posch theory, but with little mention of bivariate copula and MI. Jenison and Reale^[9] discussed how to model neural populations with copula. This paper shows that MI is actually a kind of entropy, defined as the copula entropy. A method developed to estimate the MI is then easier to use than the previous methods.

Received: 2009-03-30; revised: 2010-12-10

** To whom correspondence should be addressed.

E-mail: szq-dcs@tsinghua.edu.cn; Tel: 86-10-62788939

1 MI and Copula Entropy

Before analyzing the relation between MI and copula, a definition of the copula entropy is introduced.

Definition 1 Copula entropy Let $\mathbf{x} \in \mathbf{R}^N$ be random variables with marginal functions $\mathbf{u} = [F_1, \dots, F_N]$ and copula density $c(\mathbf{u})$. Copula entropy of \mathbf{x} is defined as:

$$H_c(\mathbf{x}) = - \int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u} \quad (3)$$

where $c(\mathbf{u}) = \frac{d^N C(\mathbf{u})}{du_1 du_2 \dots du_N}$.

As contrast, the margin entropy is defined as follows.

Definition 2 Margin entropy Let $\mathbf{x} \in \mathbf{R}^N$ be random variables with marginal functions $[p_1, \dots, p_N]$. The margin entropy is defined as

$$H(x_i) = - \int_{x_i} p_i(x_i) \log p_i(x_i) dx_i \quad (4)$$

where $i = 1, \dots, N$.

Theorem 1 The mutual information of random variables is equivalent to their negative copula entropy:

$$I(\mathbf{x}) = -H_c(\mathbf{x}) \quad (5)$$

Proof

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

Theorem 1 can be interpreted as MI exactly measuring the dependence uncertainty of random variables. This provides a new way of understanding MI as shown in Fig. 1. As defined in Eq. (1), MI used to be understood as the intersection of margin entropies. However, in Eq. (5), MI as copula entropy has no intersection with the margin entropy of each random variable because the independence between copula and margins means that (1) the variations of random variables can have different margin entropies that have no effect on the copula entropy and that (2) the same random variables can have the same margin entropy, but different types of copula functions, i.e., different types of dependent relationships, hence, different copula entropies.

The following corollary shows the relationship

between the entropy of \mathbf{x} , the margin entropy, and the corresponding copula entropy.

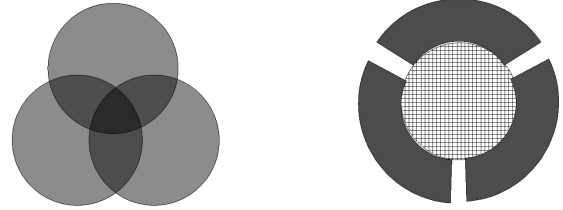


Fig. 1 A new way of understanding mutual information. Left: the traditional way of understanding MI^[2]. Right: the representation of MI as copula entropy with the copula entropy at the center surrounded by margin entropies.

Corollary 1 The entropy of joint random variables is composed of margin entropies and the copula entropy.

$$H(\mathbf{x}) = \sum_{i=1}^N H(x_i) + H_c(\mathbf{x}) \quad (6)$$

Proof

$$\begin{aligned} I(\mathbf{x}) &= \sum_i H(x_i) - H(\mathbf{x}) \Leftrightarrow H(\mathbf{x}) = \\ &= \sum_i H(x_i) - I(\mathbf{x}) = \sum_i H(x_i) + H_c(\mathbf{x}). \end{aligned}$$

Corollary 1 indicates that the entropy of random variables is composed of N margin entropies and the copula entropy, i.e., entropy of their interrelations. This relationship is illustrated in Fig. 2 which gives insight into the inner relationship between MI and copula as a connection between information theory and copula theory.

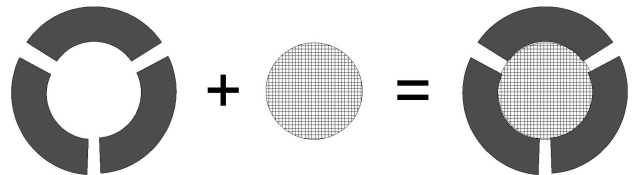


Fig. 2 The relationship between the entropy of random variables, margin entropies, and the copula entropy

2 Estimating Mutual Information via Copula

MI estimates are a common problem in many scientific fields. All traditional methods are based on Eq. (1) which leads to analyses of both margin entropy and the entropy of all the variables^[10,11]. Theorem 1 shows that the entropy estimation and the MI will be the same if the copula can be estimated. Such estimates can be achieved with empirical copula.

Given independent identical distribution samples $\mathbf{X}_T = \{X_1, \dots, X_T\}$ generated from an N dimensional $\mathbf{x} \in \mathbf{R}^N$, the MI, $I(\mathbf{x})$, can be estimated as:

Input: $\mathbf{X}_T = \{X_1, \dots, X_T\}$

Output: $I(\mathbf{x})$

Step 1 Determine the empirical copula density \hat{U}_T from \mathbf{X}_T .

Step 2 Estimate the entropy of \hat{U}_T as $I(\mathbf{x})$.

The empirical copula density, \hat{U}_T , can be determined using the empirical functions:

$$U_i^i = \frac{1}{T} \sum_{i=1}^T \mathbf{1}_{\{X_i^i \leq x_i^i\}} \quad (7)$$

where $i=1, \dots, N$ and $\mathbf{1}$ is an indicator function. Step 1 can use other methods, such as analytical or kernel methods^[12]. The empirical method is the easiest option because it is intuitive and easily calculated, since it can be solved as sorting problem. The selection of an estimation method depends on many factors, such as the background information, accuracy requirements, or personal preferences, etc. There have been many methods developed for entropy estimates, so the entropy of \hat{U}_T in Step 2 can be estimated using many well-established methods^[10,11,13].

The present method is non-parametric so all previous methods for estimating the MI can be transformed to this method which only needs the entropy estimate. Thus, the MI estimates can be achieved by simply estimating one entropy instead of many as in Eq. (1). This improvement is very significant since MI estimates are both widely-used and computationally burdensome.

3 Simulations

The effectiveness of the estimation method was evaluated using a typical case with two correlated standard Gaussian variables with covariance ρ , for which the MI can be calculated as $-\frac{1}{2} \log(1 - \rho^2)$. The tests used ρ from 0.0 to 0.9 with steps of 0.1 with a 1000 sample set generated for each value. The copula entropy was estimated using the KNN methods^[10]. The MI estimation method in Kraskov et al.^[10] was also used on the sample sets. The results in Fig. 3 show that both methods give very good estimates. The present method provides a competitive estimate of MI that is much less computationally intensive.

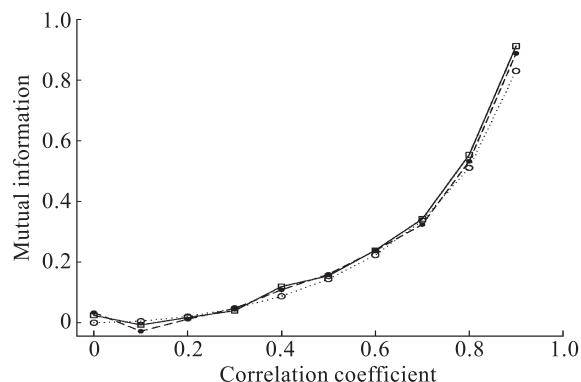


Fig. 3 Mutual information estimates. The solid line represents the analytical value of MI between two Gaussians, the dotted line represents the methods in Ref. [10], and the dashed line represents the current method.

4 Conclusions

This paper proves that MI is essentially a kind of entropy, named the copula entropy, to provide a new way of understanding MI with the theoretical relationship between information theory and copula theory. A non-parametric method is then developed to estimate the MI, which is simple and less computationally burdensome than all the previous methods based on the classical understanding of MI. This simple MI method is very useful since MI is a core concept in information theory and estimates of the MI are of fundamental importance in many applications of information theory.

References

- [1] Shannon C. A mathematical theory of communication. *Bell System Technical Journal*, 1948, **27**: 379-423.
- [2] Cover T M, Thomas J A. *Elements of Information Theory*. New York: Wiley, 1991.
- [3] Kendall M, Stuart A. *The Advanced Theory of Statistics*. Ch. Griffin, 1969.
- [4] Nelsen R B. *An Introduction to Copulas*. New York: Springer, 1999.
- [5] Joe H. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall, 1997.
- [6] Sklar A. Fonctions de repartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 1959, **8**: 229-231.
- [7] De la Peña V, Ibragimov R, Sharakhmetov S. *Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series*. Lecture Notes-Monograph Series, Institute of Mathematical Statistics, USA, 2006: 183-209.

- [8] Davy M, Doucet A. Copulas: A new insight into positive time-frequency distributions. *IEEE Signal Processing Letters*, 2003, **10**: 215-218.
- [9] Jenison R, Reale R. The shape of neural dependence. *Neural Computation*, 2004, **16**: 665-672.
- [10] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Physical Review E*, 2004, **69**: 066138.
- [11] Paninski L. Estimation of entropy and mutual information. *Neural Computation*, 2003, **15**: 1191-1253.
- [12] Silverman B. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1986.
- [13] Beirlant J, Dudewicz E, Gyoerfi L, et al. Nonparametric entropy estimation: An overview. *International Journal of Mathematical and Statistical Sciences*, 1997, **6**: 17-40.