

Transfer Entropy Estimation via Copula

Tian Xue^{1, a}

¹ Faculty of Science, Ghent University, Krijgslaan 281, 9000 Gent, Belgium

^a tian.xue@ugent.be

Keywords: Time series analysis, Transfer entropy, Copula

Abstract.

Transfer entropy provides a powerful information theoretic measurement of directed information flow between time series variables. Effective and convenient methods of estimation are desirable in practice. This article discusses the formulation of how to estimate transfer entropy via the statistical copula. Furthermore, this article provides theoretical justifications, and two estimation approaches via the Gaussian copula transformation and kernel methods. The experiment demonstrates that the proposed estimation approaches are competitive with the Linear estimator and the Nearest Neighbour estimator.

Introduction

Transfer entropy is an information-theoretic quantity that measures a directed information transfer from realizations x_t of a source stochastic processes X to the realizations y_t of target Y . Specifically, the local transfer entropy can be defined as a Shannon's local relative entropy of a q-lagged source state x_{t-q} with respect to a target state y_t given its own history y_{t-q} in time [1]:

$$\tau_{X \rightarrow Y}(t) = \log \frac{p(y_t / y_{t-q}, x_{t-q})}{p(y_t / y_{t-q})} \quad (1)$$

where $p(\cdot)$ and $p(\cdot | \cdot)$ denote the joint (conditional) probabilities of the specific realizations, and they can be modelled as the probability density function in practical. Transfer entropy $T_{X \rightarrow Y}$ is the expected value of the local transfer entropy estimates weighted by the joint probabilities of each corresponding state [2].

$$T_{X \rightarrow Y} = E\{\tau_{X \rightarrow Y}(t)\} = \sum_{y_t, y_{t-q}, x_{t-q}} p(y_t, y_{t-q}, x_{t-q}) \log \frac{p(y_t / y_{t-q}, x_{t-q})}{p(y_t / y_{t-q})} \quad (2)$$

A useful measure of $T_{X \rightarrow Y}$ shall yield a positive value in the direction of $X \rightarrow Y$ if Y depends on the past of X , while it is expected to be a zero amount of information transferred in the direction of $X \rightarrow Y$ if process Y is directed by its own past. Transfer entropy is intrinsically different from the mutual information [1]. $T_{X \rightarrow Y}$ can detect and quantify possible dependencies within any subsets of $\{x_{t-k}, y_t : t \in T, k \in Q\}$, while the mutual information only estimates an instantaneous and unidirectional information flow between two variables. As an instance, x_{t-q} may have a q-lagged shared information with y_t . Furthermore, $T_{X \rightarrow Y}$ accounts for the shared information between Y and its own past when evaluating the directed information flow from the past X to the current Y , which gives a directional sense.

Information theoretic measurement like transfer entropy has been intensively used in neuroscience such as decoding electroencephalogram (EEG) signals and constructing neural network of functional magnetic resonance imaging (fMRI) signals. However, the estimation of transfer entropy is somewhat non-trivial. There are two popular estimation approaches: the Linear estimator and the Nearest Neighbour method. The linear estimator produces efficient and consistent

results only if both process X and Y follow the Gaussian distribution. In other words, the Linear estimator requires each marginal distribution in $T_{X \rightarrow Y}$ to follow the Gaussian distribution. In practice, neuroimaging or electronic time series usually does not follow the Gaussian distribution or any other forms of frequency distribution, which biases the estimation results. The k-Nearest Neighbour estimator provides a powerful non-parametric algorithm in estimating information-theoretic quantity, but it is computationally intensive and susceptible to the number of neighbour k . Although a smaller value of k is suggested in the literature [3], it largely increases the computational burden and may result in an under-smoothed estimate. The choosing of k becomes a trade-off between the bias and variance. Therefore, it is necessary to have an easy-to-use and efficient estimation method. Inspired by the recent success of copula-based mutual information [4], a copula-based transfer entropy is naturally the next step in the field of time series analysis in information theory.

Copula-based transfer entropy

Concerning the Sklar's theorem [5], a statistical copula is a multivariate cumulative distribution function (CDF) that models the dependence structures among univariate marginal distributions. Some parametric copulas have parameters that control the strength of dependence, like the Gaussian copula. It holds further that any multivariate density function can be rewritten as a product of univariate marginal density functions and an appropriate copula density function. A conditional joint probability density $p(y_t, x_{t-q} / y_{t-q})$ can be particularly rewritten in terms of a conditional copula density function c as follows:

$$p(y_t, x_{t-q} / y_{t-q}) = p(y_t / y_{t-q}) \times p(x_{t-q} / y_{t-q}) \times c(u, v / y_{t-q}) \quad (3)$$

where $u = P(y_t / y_{t-q})$ and $v = P(x_{t-q} / y_{t-q})$ are the marginal CDFs conditional on y_{t-q} , and they are uniformly distributed in the interval $[0, 1]$. Furthermore, rearranging Eq. 3 using the conditional probability rules, and taking the natural logarithm of both sides expresses the local transfer entropy in terms of the conditional copula density function:

$$\tau_{X \rightarrow Y}(t) = \log c(u, v / y_{t-q}) = \log \frac{p(y_t / y_{t-q}, x_{t-q})}{p(y_t / y_{t-q})} \quad (4)$$

The copula-based transfer entropy thus can be obtained by evaluating its expectation over the above sample estimates.

$$T_{X \rightarrow Y} = E \left\{ \log c \left(P \left(y_t / y_{t-q} \right), P \left(x_{t-q} / y_{t-q} \right) / y_{t-q} \right) \right\} \quad (5)$$

An important corollary is that the copula-based transfer entropy does not depend on the specific marginal distribution of X and Y , which allows to estimate this novel estimator using either a known copula distribution or a non-parametric approach. The information flow that is quantified by transfer entropy will be unchanged if the empirical copula linking the variables is preserved. Therefore, this article considers two different estimation techniques: the Gaussian copula transformation and the non-parametric kernel method.

The Gaussian copula function models the dependence structure among the correlated standard Gaussian variables. As a corollary, the probability integral transformation can be used to convert the variables with arbitrary continuous distributions into standard Gaussian variables, which makes the Linear estimator feasible. In practical, empirical copulas (or pseudo observations of a statistical copula) can be easily computed by the rank normalization (the rank divided by the total number of observations). This procedure replaces original observation by its fractional rank and gives uniform

variables in the interval $[0, 1]$. Furthermore, the inverse standard Gaussian CDF is then used to evaluate these values, generating the empirical observations of the Gaussian copula function.

The non-parametric kernel method consists of a two-stage Nadaraya–Watson kernel regression [6]. The kernel regression first estimates the $P(y_t | y_{t-q})$ and $P(x_{t-q} | y_{t-q})$. The estimated conditional CDFs can be then plugged into the kernel of conditional copula density $c(u, v | y_{t-q})$. The Bernstein's approximation can be further applied, so that the estimated copulas are within the $[0, 1]$ boundary. Finally, the copula-based transfer entropy can be obtained by taking the logarithm and averaging them over the samples. Although this approach is readily straightforward, it heavily depends on the selection of kernel function and its bandwidth. An optimal bandwidth and suitable kernel function are relatively difficult to find in practice. A recent application of this approach can be found in [7].

Experiments and results:

The experiments consider two different systems of stochastic processes with a length of 1000. The purpose is to examine the numerical equivalence of copula-based transfer entropy to the other two estimation methods. The transfer entropy is then estimated by 1) the Linear estimator, 2) the k Nearest Neighbour estimator where k is set to be 1, 3) the Gaussian copula transformation and 4) the non-parametric kernel methods where the kernel function is Gaussian and the bandwidth is chosen to be 0.15 as suggested in [6, 7].

Both experiments vary the coupling strength s from 0.04 to 0.60. Results are computed as the mean of 100 runs. The first simulation experiment considers a following system:

$$\begin{cases} x_t = 0.1x_{t-1} + \delta_1 \\ y_t = 0.1y_{t-1} + sx_{t-1} + \delta_2 \end{cases} \quad (6)$$

where $(\delta_1, \delta_2) \sim N(0, \Sigma)$ and $\Sigma = 1.5$.

The Fig. 1 shows that all four methods have equivalent results. The estimation via the Gaussian copula transformation is almost matching with the Linear estimator since the underlying system is governed by the Gaussian distribution. The non-parametric kernel estimation of copula-based transfer entropy is also closed to the benchmark answers, only showing slight higher values. The Nearest Neighbour estimator fluctuates around the other three estimates, because a smaller value of k guarantees a small bias but introduces a relatively larger variance. Overall, either of these four methods promises equivalent results as long as the Gaussian distribution dominates the system. In this case, there is specific difference among these four approaches.

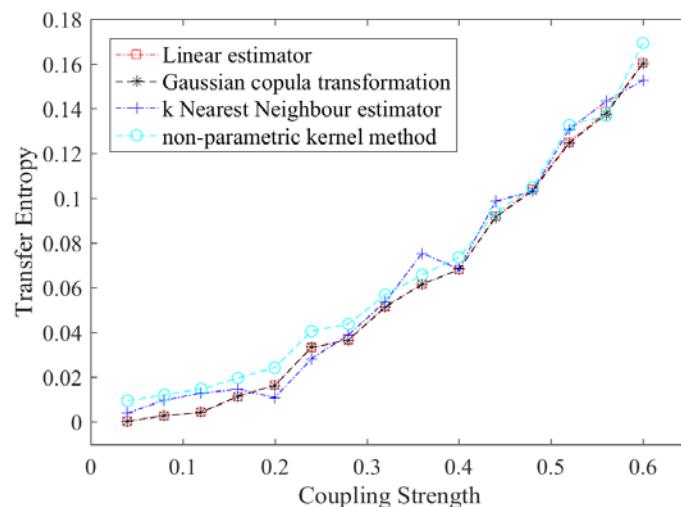


Fig. 1 Transfer entropy estimated as a function of coupling strength for Eq. 6.

In the second experiment, a more challenging task is to consider the situation where the signals are governed by a non-Gaussian distribution, for example, a Student's T distribution,

$$\begin{cases} x_t = 0.1x_{t-1} + \xi_1 \\ y_t = 0.1y_{t-1} + sx_{t-1} + \xi_2 \end{cases} \quad (7)$$

where $(\xi_1, \xi_2) \sim t_{v=1}$. Alternatively, a Student's T distribution with one degree freedom is also equivalent to a Cauchy distribution.

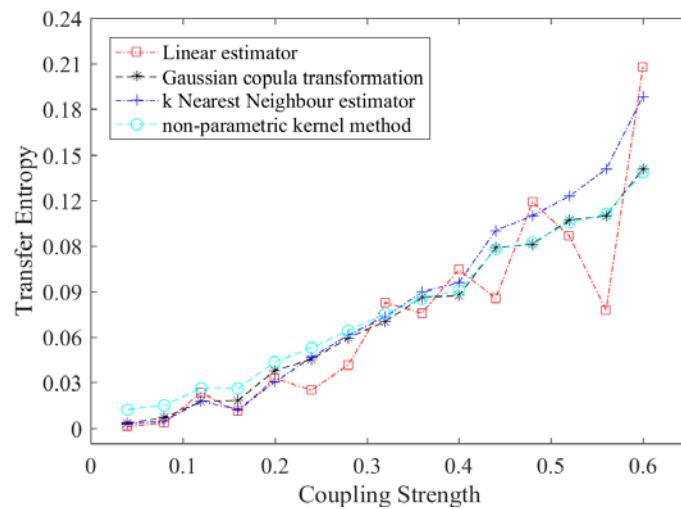


Fig. 2 Transfer entropy estimated as a function of coupling strength Eq. 7.

Based on the Fig. 2, it is obvious that the Linear estimator is quite spiky and unstable, making the estimation results unreliable. However, both estimation methods of copula-based transfer entropy provide very similar and stable results. They are directly comparable to the powerful Nearest Neighbour estimator. Nevertheless, the Gaussian copula-based transfer entropy only needs simple probability integral transformation and the Linear estimator, and both the kernel method and the Nearest Neighbour estimator require complicated specifications like the choice of Neighbour k , or the selection of kernel function and bandwidth. It means that the Gaussian copula-based transfer entropy can still be used to estimate the transfer entropy even though the underlying system is non-Gaussian. Its non-parametric counterparts can also be used if there is a careful selection rule of kernel function and its bandwidth. Therefore, the copula-based transfer entropy proposed in this letter provides desirable methods to deal with systems dominated by arbitrary marginal distributions.

Conclusion:

The development of information theoretic measures successfully provides a connection between transfer entropy and statistical copula density. The copula-based transfer entropy proposed in this letter is shown to be effective, because both estimation methods using either the Gaussian copula transformation or the kernel method provides numerical equivalence. The Gaussian copula transformation especially makes the estimation of transfer entropy simple and efficient. A statistical copula can readily measure the dependent structure among multiple variables. One advantage is that copula-based transfer entropy is unaffected by the marginal distributions of the individual variables, because the copula summarises the relationship between the two variables. A transformation method is used to find the empirical copula observations via a known copula density function, for example, Gaussian copula, such that it is feasible to use parametric linear Gaussian model for estimating. However, a direct estimation of statistical copula is computationally extensive. Note that the kernel method is powerful and featuring the non-parametric property. It is still crucial to

develop optimal selection rules for the kernel function and the bandwidth, so that computing transfer entropy is going to be more robust. This can be the direction for future researches.

References:

- [1] T. Schreiber, “Measuring information transfer,” *Phys. Rev. Lett.*, Vol. 85, 2, (2000), p. 461–464.
- [2] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, “Local information transfer as a spatiotemporal filter for complex systems,” *Phys. Rev. E*, Vol. 77, 2, (2008), p. 1–11.
- [3] A. Kraskov, H. Stögbauer, and P. Grassberger, “Estimating mutual information,” *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, Vol. 69, 6, (2004), p. 1–16.
- [4] X. Zeng and T. S. Durrani, “Estimation of mutual information using copula density function,” *Electron. Lett.*, Vol. 47, 8, (2011), p. 493–494.
- [5] A. Sklar, “Random variables, joint distribution functions, and copulas,” *Kybernetika*, Vol. 9, 6, (1973), p. 449–460.
- [6] I. Gijbels, N. Veraverbeke, and M. Omelka, “Conditional copulas, association measures and their applications,” *Comput. Stat. Data Anal.*, Vol. 55, 5, (2011), p. 1919–1932.
- [7] M. Hu and H. Liang, “A copula approach to assessing Granger causality,” *Neuroimage*, Vol. 100, ,(2014), p. 125–134.