British Journal of Mathematics & Computer Science
17(5): 1-20, 2016, Article no.BJMCS.27377
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Evaluating Transfer Entropy for Normal and γ-Order Normal Distributions

K. Hlaváčková-Schindler[1], T. L. Toulias[2*] and C. P. Kitsos[3]

[1]*Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic.*
[2]*Boulevard du Souverain 292, Auderghem 1160, Brussels, Belgium.*
[3]*Department of Informatics, Technological Educational Institute of Athens, Athens, Greece.*

Authors' contributions

This work was carried out in collaboration between all authors. Author KHS provided the framework of causality and transfer entropy as Kullback-Leibler divergence, while author CPK provided the generalized normal distribution framework as well as the Kullback-Leibler study. Author TLT performed all the mathematical computations for the explicit forms of the transfer entropy for the normal, γ-order normal and the Laplace distributions case. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2016/27377
Editor(s):
(1) Wei-Shih Du, Department of Mathematics, National Kaohsiung Normal University, Taiwan.
Reviewers:
(1) Fan Yang, Tsinghua University, Beijing, China.
(2) Abdollahzadeh Jamalabadi, Dongguk University, South Korea.
(3) Anonymous, Instituto Politecnico Nacional, Mexico
Complete Peer review History: http://sciencedomain.org/review-history/15385

Received: 30th May 2016
Accepted: 5th July 2016
Published: 13th July 2016

Original Research Article

Abstract

Since its introduction, transfer entropy has become a popular information-theoretic tool for detecting causal inference between two discretized random processes. By means of statistical tools we evaluate the transfer entropy of stationary processes whose continuous probability distributions are known. We study transfer entropy of processes coming from the family of γ-order generalized normal distribution. Applying Kullback-Leibler divergence we provide explicit expressions of the transfer entropy for processes which are normal, as well as for processes from the class of γ-order normal distributions. The results achieved in the paper for continuous time can be applied also to the discrete time case, concretely to the time series whose underlying process distribution is from the discussed classes.

**Corresponding author: E-mail: th.toulias@gmail.com;*

Keywords: *Transfer entropy; time series; Kullback-Leibler divergence; causality; generalized normal distribution.*

2010 Mathematics Subject Classification: 60GO7, 62M05, 94A17.

1 Introduction

Transfer entropy (TE), defined in [1], and equivalently in [2] via conditional mutual information, is an information-theoretic statistic, measuring the amount of directed (time-asymmetric) transfer of information between two random processes. It is a statistic working with time series. After its introduction, transfer entropy gained a wide interest in physics, neuroscience, climatology and other scientific disciplines, see for example [3], [4], among others.

Transfer entropy has been studied in relation to Granger causality (G-causality), see for example in [5], [6]. The results in [7] are asymptotic. In this paper, we study the relationship of transfer entropy and Granger causality for discretized random processes following certain probability distributions, including the family of γ-order generalized normal distributions (briefly γ-GND) as introduced in [8] and studied in [9, 10, 11, 12], delivering certain entropy and information-theoretic results. Moreover, the γ-GND was also considered in a work by [13], and the corresponding transfer entropy was in the spirit of data analysis, as introduced in the pioneering paper of Tukey [14].

2 Causality Measures in Time Series

Causality or causal inference, can be defined in terms of an effect of interventions, giving direction to the association between two variables. This approach was studied by great philosophers and recently by statisticians, as [15] among others. Another approach, coming from econometrics, called Granger causality, introduced in [16] utilizes time series analysis of the processes. It has been argued that the concept of Granger causality belongs to a different category than those of Pearl's causal model, but these concepts are closely linked, since each relates to straightforward notions of direct causality embodied in settable systems; see for example [17].

In our paper, we deal with two temporal approaches to causality inference: an econometric one, introduced by [16], and the transfer entropy (or conditional mutual information) one, as introduced by physicists [1] and [2]. In particular, we study transfer entropy by means of statistical tools, namely by considering distributions belonging to the family of γ-order generalized normal distributions, and express the corresponding transfer entropy by means of their parameters.

2.1 Granger causality

The concept of causality based on time series was introduced by C. W. J. Granger (the 2003 Nobel prize winner in economy) in [16]. Inspired by the Wiener's work, Granger based his concept on two principles:

1. The cause precedes in time the effect;
2. The cause contains information about the effect that is unique, and is in no other variable.

In other words, the causal variable can help to forecast the effect variable; see [18] for details. It is said that for *"process X_t Granger causes another process Y_t"*, if future values of Y_t can be predicted better using the past values of X_t and Y_t rather than only past values of Y_t.

The related hypotheses testing which Granger developed by [16] is based on a linear regression model, and uses two alternative test statistics, the Granger-Sargent and the Granger-Wald test.

For our future development, we adopt and use here the notation of [5]. Let d be a positive integer number and let $\oplus$ denote the concatenation of vectors, so that for $x = (x_1,\dots,x_d)$ and $y = (y_1,\dots,y_m)$, $x \oplus y$ is the $d+m$ vector $(x_1,\dots,x_d,y_1,\dots,y_m)$. Given jointly distributed multivariate random variables (r.v.-s) X and Y, i.e. random vectors in $\mathbb{R}^d$, we denote by $\Sigma(X)$ the $d \times d$ matrix of covariances $\mathrm{Cov}(X_i, X_j)$ and by $\Sigma(X,Y)$ the $d \times m$ matrix of cross-covariances $\mathrm{Cov}(X_i, Y_k)$. Let $\Sigma(X|Y)$ denote the $d \times d$ matrix

$$\Sigma(X|Y) = \Sigma(X) - \Sigma(X,Y)\Sigma(Y)^{-1}\Sigma(X,Y)^{\mathrm{T}}, \tag{2.1}$$

whenever $\Sigma(Y)$ is invertible.

Assume a multivariate stochastic process X_t in discrete time (i.e. marginal distributions are jointly distributed). Denote with $X_t^{(k)} = X_t \oplus X_{t-1} \oplus \cdots \oplus X_{t-k+1}$ for X along with $k-1$ lags so that $X_t^{(k)}$ is a kd random vector for each t. Given the lag k, where it is clear, we use the shorthand notation $X_t^- \equiv X_{t-1}^{(k)}$ for the lagged variable.

Suppose we have three jointly distributed stationary multivariate stochastic processes X_t, Y_t, Z_t. Consider the regression models

$$X_t = \alpha_t + \left(X_{t-1}^{(k)} \oplus Z_{t-1}^{(r)}\right) A + \varepsilon_t, \tag{2.2a}$$

$$X_t = \alpha'_t + \left(X_{t-1}^{(k)} \oplus Y_{t-1}^{(q)} \oplus Z_{t-1}^{(r)}\right) A' + \varepsilon'_t, \tag{2.2b}$$

where A and A' are the matrices of regression coefficients, α_t and α'_t are the constant terms and the random vectors ε_t and ε'_t comprise the residuals, so that the predictee variable X is regressed firstly on the previous k lags of itself plus r lags of the conditioning variable Z and secondly, in addition, on q lags of the predictor variable Y.

The G-causality of Y to X given Z assesses the extent to which inclusion of Y in the second model (2.2b) reduces the prediction error of the first model (2.2a). The standard measure of G-causality in the literature is defined for the predictor and predicted variables Y and X respectively, and is given by the natural logarithm of the ratio of the residual variance in the restricted regression (2.2a) to that of the unrestricted regression (2.2b). It has been proven by [5] that for the G-causality it holds

$$\mathscr{F}_{Y\to X|Z} = \log \frac{\Sigma(X|X^- \oplus Z^-)}{\Sigma(X|X^- \oplus Y^- \oplus Z^-)}, \tag{2.3}$$

where $\mathscr{F}_{Y\to X|Z}$ denotes the G-causality for a univariate predictor and predicted variables Y and X, conditioned by Z.

2.2 Transfer entropy

This section is based on the definitions given in [19]. Transfer entropy (TE) is a model-free, information-theoretic expression measuring the amount of time-directed information between two dynamical systems, which was introduced by [1], and contemporarily by [2] as conditional mutual information; for comparison see [20]. Given the past time of a dynamical system X, then the TE from another dynamical system Y to the first system X is the amount of the reduction of Shannon uncertainty measured in the future time of X when the past of Y is given. Since its introduction it has been extensively applied in modeling complex systems, especially in neuroscience and climatology, see e.g. [4, 3]. Transfer entropy is based on Shannon entropy, which is defined for X, a discrete (multivariate) r.v. given by a set of possible values $\{x_1, x_2 \dots, x_n\}$, as

$$H(X) := \sum_{i=1}^{n} p(x_i) \ln p(x_i), \tag{2.4}$$

where p denotes the probability mass function of X. With X_t, Y_t, and Z_t as defined earlier, the transfer entropy of Y to X given Z, denoted by $\mathscr{T}_{Y\to X|Z}$, is then defined as the difference between the entropy of

X conditioned on its own past and the past of Z, and its entropy conditioned on the past of Y:

$$\mathcal{T}_{Y\to X|Z} := H(X|X^- \oplus Z^-) - H(X|X^- \oplus Y^- \oplus Z^-), \tag{2.5}$$

where $H(.|.)$ is the conditional entropy. In other words, $\mathcal{T}_{Y\to X|Z}$ denotes the transfer entropy of the time series (stochastic process) Y to the time series (stochastic process) X under the condition of time series (stochastic process) Z. For stationary variables, transfer entropy does not depend on t, so we shall exclude it from labeling.

For a broad class of predictive models, Barnett and Bossomaier showed in [7], that the log-likelihood ratio test statistic for the null hypothesis of zero transfer entropy is a consistent estimator for the transfer entropy itself. An asymptotic chi-squared distribution was established for the transfer entropy estimator. Their result generalizes the equivalence of transfer entropy and Granger causality in the Gaussian case and bridges the notion of directed information transfer and G-causality.

In addition to these general results, it has been proven that the complexity of approximation of entropy is polynomial; see e.g. [21]. However, practical methods to achieve a good approximation of (differential) entropy are non-trivial; see for details [22] or [23] among others.

3 Continuous Transfer Entropy

Zhu et al. in [24] formulated the definition of causality by transfer entropy for continuous processes via Markov processes, equivalent to transfer entropy definition as in [25]. For our computations in what follows, we adopt that definition.

Denote by $X_t^{(k)} = (X_{t-k+1}, X_{t-k+2}, \dots, X_t)$, $k > 0$, and assume the probability measure $\mathscr{P}_X$ (defined on measurable subsets of real sequences) on X fulfills the $m-$ th order Markov property:

$$\text{for all } t: \text{ for all } m' > m: \ \mathrm{d}\mathscr{P}_{X_{t+1}|X_t^{(m)}}\left(x_{t+1}|x_i^{(m)}\right) = \mathrm{d}\mathscr{P}_{X_{t+1}|X_j^{(m')}}\left(x_{t+1}|x_t^{(m')}\right),$$

for $x_{t+1} \in \mathbb{R}$, $x_t^{(k)} \in \mathbb{R}^k$. Then, the past information $X_t^{(m)}$ (preceding the time instant $t+1$) is sufficient for predicting X_{t+k}, $k \geq 1$, and can be considered as an m-dimensional state vector at time t. Assume further that N is a positive integer corresponding to the length of the discretized time series $(\hat{X}_t, X_t^-, Y_t^-)$, $t = 1, 2, \dots, N$. For independent and identically distributed (i.i.d.) random series, each term has the same distribution as the random vector $(\hat{X}, X^-, Y^-) \in \mathbb{R}^{1+m+n}$, whatever i is considered. For random variables $\hat{X}$, X^-, Y^-, note that "ˆ" indicates "predicted", while "−" means "past", which is a common notation in the literature; see for example [20] among others.

3.1 Causality defined by deviation from Markov property

In this section we use the terms random variable and random process interchangeably. Let us suppose that a causal influence exists from random process Y to process X, and is such that at each time i integer and for some $n > 0$, Y_t^- is an image of the physical state Y, and it can be written as $Y_t^- \triangleq Y_t^{(n)}$. The negation of this causal influence implies that for given t:

$$\mathrm{d}\mathscr{P}_{\hat{X}_t|X_t^{(m)}}\left(\hat{x}_t|x_t^{(m)}\right) = \mathrm{d}\mathscr{P}_{\hat{X}_t|X_t^{(m)},Y_t^{(n)}}\left(\hat{x}_t|x_t^{(m)}, y_t^{(n)}\right), \text{ for all } m, n > 0, \tag{3.1}$$

defines the causal flow. If (3.1) holds, it is said that there is an absence of information transfer from Y to X.

Otherwise, the process X can be no longer considered strictly as a Markov process. Let us suppose the joint process (X, Y) is Markovian, i.e. there exists a given pair (m', n'), a transition function f and an independent random sequence ε_t, $t \in \mathbb{Z}$ ($\mathbb{Z}$ denotes the set of integer numbers), such that $[X_{t+1}, Y_{t+1}]^{\mathrm{T}} = f\big(X_i^{(m')}, Y_t^{(n')}, \varepsilon_{t+1}\big)$, where the r.v. ε_{t+1} is independent of the past random sequence $(X_j, Y_j, \varepsilon_j)$, $j \leq t$, whatever t is considered. As $X_t = g\big(X_t^{(m)}, Y_t^{(n)}\big)$, where g is clearly a non-injective function, the triplet $\big(X_t^{(m)}, Y_t^{(n')}, X_t\big)$, $t \in \mathbb{Z}$, corresponds to a hidden Markov process, and it is well-known that this observation process is not generally Markovian, [24].

Similarly as in Section 2.2, for the discrete transfer entropy, if the processes X and Y are assumed to be jointly stationary, then the expectation of the form $\mathrm{E}\big(g\big(X_{t+1}, X_t^{(m)}, Y_t^{(n)}\big)\big)$ does not depend on t for any function $g : \mathbb{R}^{m+n+1} \to \mathbb{R}$, see [24]. This assumption is essential for the construction of transfer entropy, since it allows us to omit the index t, as we are moving from the discrete to the continuous case.

3.2 Transfer entropy as Kullback-Leibler divergence

Recall the Kullback-Leibler (KL) divergence, defined by [26], is a non-symmetric measure of the difference between X and Y as

$$D_{\mathrm{KL}}(X \| Y) := \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{q(x)} \,\mathrm{d}x, \tag{3.2}$$

where $p(x)$ and $q(x)$, $x \in \mathbb{R}^d$ being the probability densities of X and Y respectively.

Formulated using probability theory, transfer entropy is a conditional mutual information, [2], [22].

The deviation from Markov property can be expressed through the Kullback-Leibler divergence which leads to the definition of transfer entropy at time t:

$$TE_{Y \to X;t} = \int_{\mathbb{R}^{1+m+n}} \log \left\{ \frac{\mathrm{d}\mathscr{P}_{\hat{X}_t | X_t^-, Y_t^-}\big(\hat{x}_t | x_t^-, y_t^-\big)}{\mathrm{d}\mathscr{P}_{\hat{X}_t | X_t^-}\big(\hat{x}_t | x_t^-\big)} \right\} \mathrm{d}\mathscr{P}_{\hat{X}_t | X_t^-, Y_t^-}\big(\hat{x}_t | x_t^-, y_t^-\big). \tag{3.3}$$

If the joint probability measure $\mathscr{P}_{\hat{X}_t | X_t^-, Y_t^-}(\hat{x}_t | x_t^-, y_t^-)$ can be derived with respect to the Lebesgue measure μ in $\mathbb{R}^{1+m+n}$, then the (joint) probability density function (p.d.f.) exists (and also each p.d.f. for each subset of $\hat{X}_t | X_t^-, Y_t^-$); we denote it as $p_{\hat{X}_t | X_t^-, Y_t^-}(\hat{x}_t | x_t^-, y_t^-)$. Then, transfer entropy $TE_{Y \to X;t}$ can be rewritten as

$$TE_{Y \to X;t} = \int_{\mathbb{R}^{1+m+n}} p_{\hat{X}, X_t^-, Y_t^-}(\hat{x}, x_t^-, y_t^-) \log \frac{p_{\hat{X}, X_t^-, Y_t^-}\big(\hat{x}, x_t^-, y_t^-\big) p_{X_t^-}\big(x_t^-\big)}{p_{X_t^-, Y_t^-}\big(x_t^-, y_t^-\big) p_{\hat{X}, X_t^-}\big(\hat{x}_t, x_t^-\big)} \,\mathrm{d}\hat{x}_t \,\mathrm{d}x_t^- \,\mathrm{d}y_t^-. \tag{3.4}$$

More details to the derivation of the definition of transfer entropy can be found in [24].

Now we proceed as follows. Denote $u := \hat{x}_t$, $v := x_t^-$ and $w := y_t^-$ (assuming stationary processes), where $U := \hat{X}$, $V := X_t^-$, $W := Y_t^-$. We can define a simpler form of (3.4) through the definition of the triple (u, v, w) and the corresponding probabilistic relationship:

$$q(u, v, w) := \frac{p(v, w) p(u, v)}{p(v)}. \tag{3.5}$$

Then through (3.5) and (3.4) we obtain the more simplified form,

$$TE_{Y \to X;t} = TE_{W \to U|V} = \int_{\mathbb{R}^{1+m+n}} p(u, v, w) \log \frac{p(u, v, w) p(v)}{p(v, w) p(u, v)} \,\mathrm{d}u \,\mathrm{d}v \,\mathrm{d}w \tag{3.6}$$

$$= \int_{\mathbb{R}^{1+m+n}} p(u, v, w) \log \frac{p(u, v, w)}{q(u, v, w)} \,\mathrm{d}u \,\mathrm{d}v \,\mathrm{d}w, \tag{3.7}$$

where p and q are probability distributions.

Equation (3.7) for TE will be investigated as the Kullback-Leibler measure between two probability distributions. In this paper, we assume that $p(\cdot)$ is a γ-order generalized normal p.d.f. (p follows the γ-GND). These will be briefly discussed in the following section.

The causality, measured by transfer entropy in (3.4) and, in agreement with notation in Section 3 can be tested with hypothesis $\mathrm{H}_0 : U \perp W|V$, where $\perp$ denotes probabilistic independence.

4 Family of γ-Order Generalized Normal Distributions

[8] came with the idea of generalizing Fisher's entropy type information, the entropy type measure from

$$J(X) := \int_{\mathbb{R}^d} \left|\nabla \log f\right|^2 f \,\mathrm{d}x = \int_{\mathbb{R}^d} \nabla f \cdot \nabla \log f \,\mathrm{d}x, \tag{4.1}$$

see [27], to

$$J_\alpha(X) := \int_{\mathbb{R}^d} \left|\nabla \log f\right|^\alpha f \,\mathrm{d}x = \int_{\mathbb{R}^d} \left|\nabla f\right|^\alpha \cdot \nabla \log f^{1-\alpha} \,\mathrm{d}x, \tag{4.2}$$

for r.v. X with probability density f on $\mathbb{R}^d$.

Applying the optimal Euclidean logarithm Sobolev inequality, [8] noted that its extremals (i.e. the function that makes the Sobolev inequality to be an equality) offered a generalized form of the multivariate normal distribution, called as the γ-order normal distribution, and denoted with $\mathcal{N}_\gamma(\mu, \Sigma)$. The density function of a r.v. $X \sim \mathcal{N}_\gamma^d(\mu, \Sigma)$ is of the form

$$f_X(x;\, \mu, \Sigma) = C(\gamma, d)|\Sigma|^{-1/2} \exp\left\{-\tfrac{\gamma-1}{\gamma} Q_X(x)^{\frac{\gamma}{2(\gamma-1)}}\right\}, \quad x \in \mathbb{R}^d, \tag{4.3}$$

where $Q_X(x) := \langle x-\mu, \Sigma^{-1}(x-\mu)^{\mathrm{T}}\rangle = (x-\mu)\Sigma^{-1}(x-\mu)^{\mathrm{T}}$, $|\Sigma|$ denotes the determinant of Σ, and the normalizing factor, for $\gamma \in \mathbb{R} \setminus [0,1]$ is given by

$$C^d = C(\gamma, d) := \pi^{-\frac{d}{2}} \frac{\Gamma\left(\frac{d}{2}+1\right)}{\Gamma\left(d\frac{\gamma-1}{\gamma}+1\right)} \left(\tfrac{\gamma-1}{\gamma}\right)^{d\frac{\gamma-1}{\gamma}}. \tag{4.4}$$

For details see Kitsos and Tavoularis in [8].

The generalized Fisher entropy type information, as well as the generalized entropy power for a r.v. $X \sim \mathcal{N}_\gamma^d(\mu, \Sigma)$ were studied in [9] and [28]. Moreover, the maximum likelihood estimation (MLE) for the γ-GND family is studied in [11]. The family of γ-order GND is based on a solid theoretical background and extensively studied in [10], [11], and [12], for the multivariate case. Notice that with $\gamma = 2$ we get the well known $\mathcal{N}_2^d(\mu, \Sigma)$.

One of the merits of the γ-order GND is that "fat tails" can be produced for a particular γ value. The shape parameter γ influences the "amount of probability at the tails" of the normal distribution. Moreover, from an information-theoretic point of view, the γ-GND provides the equality in the generalized Cramér-Rao inequality, introduced by [8]), just like the usual normal distribution does for the usual Cramér-Rao inequality; see also [29].

The γ-GND is an elliptically contoured distribution when its scale matrix Σ is a positive definite matrix, while is spherically contoured when $\Sigma := \sigma^2 \mathbb{I}_p$.

The family of $\mathcal{N}_\gamma^d(\mu, \Sigma)$, i.e. the family of the elliptically contoured γ-order normals, provides a "smooth bridging" between the multivariate (and elliptically countered) uniform, normal and Laplace r.v.-s U, N and L respectively, i.e. between $U \sim \mathscr{U}^d(\mu,\Sigma)$, $N \sim \mathscr{N}^d(\mu,\Sigma)$ and Laplace $L \sim \mathscr{L}^d(\mu,\Sigma)$ r.v.-s as well as the multivariate degenerate Dirac distributed r.v. $D \sim \mathscr{D}^d(\mu)$, with pole at the point μ. Recall the p.d.f-s of U, N, L, and D, i.e.

$$f_U(x) = f_U(x;\,\mu,\Sigma) := \begin{cases} \dfrac{\Gamma\left(\frac{d}{2}+1\right)}{\pi^{d/2}\sqrt{|\Sigma|}}, & x \in A_\theta, \\ 0, & x \notin A_\theta, \end{cases} \tag{4.5a}$$

$$f_N(x) = f_N(x;\,\mu,\Sigma) := \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left\{-\tfrac{1}{2}Q_\theta(x)\right\}, \quad x \in \mathbb{R}^d, \tag{4.5b}$$

$$f_L(x) = f_L(x;\,\mu,\Sigma) := \frac{\Gamma\left(\frac{d}{2}+1\right)}{d!\pi^{d/2}\sqrt{|\Sigma|}} \exp\left\{-\sqrt{Q_\theta(x)}\right\}, \quad x \in \mathbb{R}^d, \tag{4.5c}$$

$$f_D(x) = f_D(x;\,\mu) := \begin{cases} +\infty, & x = \mu, \\ 0, & x \in \mathbb{R}^d \setminus \{\mu\}, \end{cases} \tag{4.5d}$$

where A_θ denotes the area enclosed by the d-ellipsoid defined by Q_θ, i.e. $A_\theta :\ Q_\theta(x) \leq 1,\ x \in \mathbb{R}^p$, with $Q_\theta(x) =: (x-\mu)\Sigma^{-1}(x-\mu)^{\mathrm{T}}$, $\theta := (\mu,\Sigma)$.

That is, the $\mathscr{N}_\gamma^d$, family of distributions not only contains the usual normal, but two other very significant distributions, as the uniform and Laplace distributions, are also members of this family, together with the degenerated Dirac distribution. The above discussion is summarized in the following theorem; see also [10].

Theorem 4.1. *The elliptically contoured d-variate γ-order normal distribution $\mathscr{N}_\gamma^d(\mu,\Sigma)$ for order values of $\gamma = 0, 1, 2, \pm\infty$ coincides with*

$$\mathscr{N}_\gamma^d(\mu,\Sigma) = \begin{cases} \mathscr{D}^d(\mu), & \textit{for } \gamma = 0 \textit{ and } d = 1,2, \\ 0, & \textit{for } \gamma = 0 \textit{ and } d \geq 3, \\ \mathscr{U}^d(\mu,\Sigma), & \textit{for } \gamma = 1, \\ \mathscr{N}^d(\mu,\Sigma), & \textit{for } \gamma = 2, \\ \mathscr{L}^d(\mu,\Sigma), & \textit{for } \gamma = \pm\infty. \end{cases} \tag{4.6}$$

The above briefly discussed properties of the γ-GND are applied in Section 6, generalizing the results that we obtained for normal distribution in Section 5. Moreover, Section 7 develops the achieved results and provides examples of the known distributions this family includes. For different γ values (shape parameter) close to the value 2, different "normal-like distributions" can be obtained.

5 Explicit Form of Transfer Entropy for the Normal Distribution

The joint probability density $p_{\hat{X}_i|X_i^-,Y_i^-}(\hat{x}_i|x_i^-,y_i^-)$, as in (3.4), is considered now to be a d-variate normal distribution, with $d := 1+m+n$, $m,n \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$, $m \neq n$. In order to calculate $TE_{W\to U|V}$ as in (3.6), first we consider a vector $x \in \mathbb{R}^d$ such that $x = (x_i)_{i=1}^d := u \oplus v \oplus w$, where $u := x_1 \in \mathbb{R}$, $v := (x_2, x_3, \ldots, x_{m+1}) \in \mathbb{R}^m$, and $w := (x_{m+2}, x_{m+3}, \ldots, x_{d=1+m+n}) \in \mathbb{R}^n$. In general, for every d-dimensional vector $x \in \mathbb{R}^d$, we can write $x = x_u \oplus x_v \oplus x_w$, where vectors x_u, x_v and x_w are the corresponding 1-, m- and n-dimensional components of $x \in \mathbb{R}^d$, like u, v, and w as above. Moreover, let $x_{u\oplus v} := x_u \oplus x_v = x_1 \oplus x_v \in \mathbb{R}^{m+1}$ and $x_{v\oplus w} := x_v \oplus x_w \in \mathbb{R}^{d-1} = \mathbb{R}^{m+n}$.

Consider now a d-variate and spherically contoured normally distributed r.v., say X, i.e. $X \sim \mathscr{N}^d\left(\mu, \sigma^2\mathbb{I}_d\right)$

with mean vector $\mu = (\mu_i)_{i=1}^d \in \mathbb{R}^d$ and $\sigma > 0$ with $\mathbb{I}_d \in \mathbb{R}^{d\times d}$ denoting the identity $d \times d$ matrix. Using the notation as above, we can adopt the m-variate, $(m+1)$-variate, and $(m+n)$-variate normally distributed random variables $X_v \sim \mathcal{N}^m(\mu_v, \sigma^2\mathbb{I}_m)$, $X_{u\oplus v} \sim \mathcal{N}^{1+m}(\mu_{u\oplus v}, \sigma^2\mathbb{I}_{1+m})$, as well as $X_{v,w} \sim \mathcal{N}^{m+n}(\mu_{v\oplus w}, \sigma^2\mathbb{I}_{m+n})$.

Assume now that the joint probability density $p(u,v,w)$ of the transfer entropy $W \to U|V$, as in the simplified form of (3.6), corresponds to the $(1+m+n)$-variate normal distribution followed by r.v. X as above, i.e. $p(u,v,w) := p_X(x)$, $x \in \mathbb{R}^d$, with random variable $X := U \oplus V \oplus W \sim \mathcal{N}^d(\mu, \sigma^2\mathbb{I}_p)$. Then the p.d.f.-s $p(v)$, $p(u,v)$ and $p(v,w)$ can be calculated and hence the value of $TE_{W\to U|V}$ can be obtained. Indeed, we consider

$$p(u,v,w) := p_X(x) = \left(\sqrt{2\pi}\sigma\right)^{-d} \exp\left\{-\tfrac{1}{2}\left\|\frac{x-\mu}{\sigma}\right\|^2\right\}, \quad x \in \mathbb{R}^d, \tag{5.1}$$

where $\|\cdot\|$ denotes the $\mathcal{L}^2(\mathbb{R}^d)$ norm.

Thus, we let $p(u) := p_U(u)$, $p(u,v) := p_{U\oplus V}(u,v)$, $p(v,w) := p_{V\oplus W}(v,w)$ and $p(u,v,w) := p_{U\oplus V\oplus W}(u,v,w)$, where the random variables $V := X_v$, $U \oplus V := X_{u\oplus v}$ and $V \oplus W := X_{v\oplus w}$. Therefore,

$$TE_{W\to U|V} = \int_{\mathbb{R}^d} p_{U\oplus V\oplus W} \log \frac{p_{U\oplus V\oplus W}\, p_V}{p_{U\oplus V}\, p_{V\oplus W}}, \tag{5.2}$$

which is a mathematically correct expression (using random variables) of (3.6), which shall be used here. Substituting (5.1) and $p(u)$, $p(u,v)$ and $p(v,w)$ as described above, into (5.2) we obtain, after some computations, that

$$TE_{W\to U|V} = -\frac{1}{2(2\pi)^{d/2}\sigma^d} \int_{\mathbb{R}^d} \left(\left\|\frac{x-\mu}{\sigma}\right\|^2 + \left\|\frac{x_v-\mu_v}{\sigma}\right\|^2 - \left\|\frac{x_{v\oplus w}-\mu_{v\oplus w}}{\sigma}\right\|^2\right) \exp\left\{-\tfrac{1}{2}\left\|\frac{x-\mu}{\sigma}\right\|^2\right\} \mathrm{d}x. \tag{5.3}$$

Through the linear transformation $z = z(x) := (x-\mu)/\sigma$, $x \in \mathbb{R}^d$, which implies $\mathrm{d}x = \mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_d = \sigma^d\,\mathrm{d}z_1\,\mathrm{d}z_2 \cdots \mathrm{d}z_d = \sigma^d\,\mathrm{d}z$, we have

$$TE_{W\to U|V} = -\tfrac{1}{2}(2\pi)^{-d/2} \int_{\mathbb{R}^d} \left(\|z\|^2 + \|z_v\|^2 - \|z_{u\oplus v}\|^2 - \|z_{v\oplus w}\|^2\right) \mathrm{e}^{-\frac{1}{2}\|z\|^2} dz. \tag{5.4}$$

However, as

$$\begin{aligned}
\|z\|^2 &= z_1^2 + \left(z_2^2 + z_3^2 + \cdots + z_{m+1}^2\right) + \left(z_{m+2}^2 + z_{m+3}^2 + \cdots + z_d^2\right), && \text{(5.5a)}\\
\|z_v\|^2 &= z_2^2 + z_3^2 + \cdots + z_{m+1}^2, && \text{(5.5b)}\\
\|z_{u\oplus v}\|^2 &= z_1^2 + \left(z_2^2 + z_3^2 + \cdots + z_{m+1}^2\right), && \text{(5.5c)}\\
\|z_{v\oplus w}\|^2 &= \left(z_2^2 + z_3^2 + \cdots + z_{m+1}^2\right) + \left(z_{m+2}^2 + z_{m+3}^2 + \cdots + z_d^2\right), && \text{(5.5d)}
\end{aligned}$$

it holds, eventually, that $TE_{W\to U|V} = 0$ after substituting (5.5a)–(5.5d) into (5.4). Therefore, the transfer entropy with normally distributed and spherically contoured joint distribution vanishes. Indeed, the concatenated vectors z in eqs. (5.5a)-(5.5d) imply that $TE = 0$. For the most general multivariate normal distribution, i.e. the elliptically contoured normal, can be in general TE non-zero.

On the other hand, it is well-known that the conditional mutual information $I(U,W|V)$ is zero if and only if random variables $U|V$ and $W|V$ are independent. Since we can express transfer entropy, according to [2] by means of conditional mutual information

$$TE_{W\to U|V} = I(U,W|V),$$

the statement holds also for transfer entropy.

For $TE_{W \to U|V}$ rewritten in our notation as $TE_{Y_t^- \to \hat{X}|X_t^-}$, it holds that $TE_{Y_t^- \to \hat{X}|X_t^-} = 0$ if and only if the processes $(\hat{X}|X_t^-)$ and $(Y_t^-|X_t^-)$ are independent. This is the case when Markov property (3.1) is not violated, i.e. there is (for stationary processes) no information transfer from Y to X; in other words Y_t has no causal influence on X_t.

The vanishing of $TE_{W \to U|V}$ can also be justified by the fact that (5.2) is written as a KL divergence between $p(u,v,w)$ (normally distributed) and $q(u,v,w)$ as in (3.5). This is because, substituting $p(u)$, $p(u,v)$ and $p(v,w)$, as described earlier, into (3.5), then $q(u,v,w)$ coincides with $p(u,v,w)$ (the corresponding summations inside the exponential function follow the same pattern as shown by the relations (5.5b)–(5.5d)). Thus, the KL divergence between the same p.d.f. is zero, and hence $TE_{W \to U|V} = 0$.

6 Explicit Form of Transfer Entropy for the γ-GND

Consider now that the joint probability density $p(u,v,w)$ of the transfer entropy $TE_{W \to U|V}$ in (3.6) corresponds to the $(1+m+n)$-variate γ-order normal distribution. In particular, let $X := U \oplus V \oplus W \sim \mathcal{N}_\gamma^d(\mu, \sigma^2 \mathbb{I}_d)$ with mean vector $\mu = (\mu_i)_{i=1}^q \in \mathbb{R}^d$ and $\sigma > 0$. Its p.d.f. is then given by (4.3). We also adopt the m-variate, $(m+1)$-variate, and $(m+n)$-variate γ-GND r.v.-s $V := X_v \sim \mathcal{N}_\gamma^m(\mu_v, \sigma^2 \mathbb{I}_m)$, $U \oplus V := X_{u \oplus v} \sim \mathcal{N}_\gamma^{1+m}(\mu_{u \oplus v}, \sigma^2 \mathbb{I}_{1+m})$, and $V \oplus W := X_{v \oplus w} \sim \mathcal{N}_\gamma^{m+n}(\mu_{v \oplus w}, \sigma^2 \mathbb{I}_{m+n})$, respectively.

The following Lemma is needed for the calculation of the transfer entropy for the general case of the γ-order GND. Recall that $(r)_k := r(r-1)(r-2)\cdots(r-k+1)$, $r \in \mathbb{R}$, $k \in \mathbb{N}$, denotes the Pochhammer symbol.

Lemma 6.1. *The transfer entropy $TE_{W \to U|V}$, when $U \oplus V \oplus W \sim \mathcal{N}_\gamma^d(\mu, \sigma^2 \mathbb{I}_d)$, $d := 1+m+m$, equals to*

$$TE_{W \to U|V} = \log C - \frac{C^d \sigma^d}{g} \sum_{k=0}^{\infty} \frac{(g/2)_k}{k!} \sum_{l=0}^{k} \binom{k}{l} (-1)^l I_{k-l} \sum_{T_d^{k-l}} \frac{(k-l)!}{t_1! \cdots t_d!} \prod_{\ell=1}^{d-1} J_\ell, \tag{6.1}$$

where $C := C^d C^m / (C^{1+m} C^{m+n})$, $g := \gamma/(\gamma-1)$, and T_d^{k-l} is considered to be the set of indexes $T_d^{k-l} := \{(t_1, t_2, \ldots, t_d) \in \mathbb{N}_1^d : t_1 + t_2 + \cdots + t_d = k-l$ and $(0 < t_1 < k-l$ or $0 < t_{m+2} + \cdots + t_d < k-l)\}$. Moreover,

$$I_{k-l} := \int_0^{+\infty} \rho^{2(k-l)+d-1} \exp\left\{-\tfrac{1}{g}\rho^g\right\} \mathrm{d}\rho, \quad k-l \in \mathbb{N}, \tag{6.2a}$$

$$J_\ell := \int_0^{\pi} \sin^{\tau_\ell} \varphi_\ell \cos^{2t_\ell} \varphi_\ell \, \mathrm{d}\varphi_\ell, \quad \ell = 1,2,\ldots,d-2, \tag{6.2b}$$

$$J_{d-1} := \int_0^{2\pi} \sin^{\tau_{d-1}} \varphi_{d-1} \cos^{2t_{d-1}} \varphi_{d-1} \, \mathrm{d}\varphi_{d-1}. \tag{6.2c}$$

See Appendix A for the proof.

The following theorem provides the transfer entropy for the general case of the γ-order GND, and therefore generalizes the result in Section 5.

Theorem 6.2. *The transfer entropy $TE_{W \to U|V}$ for $U \oplus V \oplus W \sim \mathcal{N}_\gamma^d(\mu, \sigma^2 \mathbb{I}_d)$, is calculated to be*

$$\begin{aligned} TE_{W \to U|V} &= \log \frac{\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{1+m}{g}\right)\Gamma\left(\frac{m+n}{g}\right)}{\Gamma\left(\frac{d}{g}\right)\Gamma\left(\frac{m}{g}\right)\Gamma\left(\frac{1+m}{2}\right)\Gamma\left(\frac{m+n}{2}\right)} - \\ &\frac{\Gamma\left(\frac{d}{2}\right)}{\pi^{d/2}\Gamma\left(\frac{d}{g}\right) g^{(d+g)/g}} \sum_{k=0}^{\infty} \frac{(g/2)_k}{k!} \sum_{l=0}^{k} \binom{k}{l} (-1)^l g^{\frac{2(k-l)+q}{g}} \Gamma\left(\frac{2(k-l)+q}{g}\right) \sum_{T_d^{k-l}} \frac{(k-l)!}{t_1! t_2! \cdots t_q!} P_d, \end{aligned} \tag{6.3}$$

where

$$P_d := 2\pi^{\frac{m+n}{2}} \left[\prod_{\ell=0}^{\frac{m+n}{2}-1} \frac{(-1)^{\frac{\tau_{2\ell+1}+1}{2}}}{2^{\tau_{2\ell+1}-2}} R_{2\ell+1} \sum_{j=0}^{(\tau_{2\ell+1}-1)/2} \frac{(-1)^j}{\tau_{2\ell+1}-2j} \binom{\tau_{2\ell+1}}{j} \right] \prod_{\ell=1}^{\frac{m+n}{2}} \frac{R_{2\ell}}{2^{\tau_{2\ell}}} \binom{\tau_{2\ell}}{\tau_{2\ell}/2}, \tag{6.4}$$

for $m,n \in \mathbb{N}^*$ being both odd or even numbers, and

$$P_d := 2\pi^{\frac{d}{2}-1} \left[\prod_{\ell=0}^{\frac{m+n-1}{2}} \frac{R_{2\ell+1}}{2^{\tau_{2\ell+1}}} \binom{\tau_{2\ell+1}}{\tau_{2\ell+1}/2} \right] \prod_{\ell=1}^{\frac{m+n-1}{2}} \frac{(-1)^{(\tau_{2\ell}-1)/2}}{2^{\tau_{2\ell}-2}} R_{2\ell} \sum_{j=0}^{(\tau_{2\ell}-1)/2} \frac{(-1)^j}{\tau_{2\ell}-2j} \binom{\tau_{2\ell}}{j}, \tag{6.5}$$

for $m,n \in \mathbb{N}^*$ not being both (simultaneously) odd or even numbers, with $\tau_\ell := 2(t_{\ell+1}+t_{\ell+2}+\cdots+t_d) + m+n-\ell$, $\ell = 1,2,\ldots,m+n$, and for $k \in \mathbb{N}$,

$$R_k := \frac{(2t_k-1)!!}{(2t_k+\tau_k)(2t_k+\tau_k-2)\cdots(\tau_k+2)} = \frac{(2t_k-1)!!}{(2t_k+\tau_k)^{t_k}\left(t_k+\frac{1}{2}\tau_k\right)_{t_k}}. \tag{6.6}$$

See Appendix B for the proof.

It is worth mentioning that the transfer entropy under the γ-GND family $\mathcal{N}_\gamma^d(\mu,\Sigma)$, $\Sigma := \sigma^2\mathbb{I}_d$, is invariant in terms of the mean $\mu \in \mathbb{R}^d$ and the (spherically contoured) scale matrix $\Sigma \in \mathbb{R}^{d\times d}$.

7 Special Cases

Theorem 6.2 can be applied to all members of the γ-GND family. The cases of normal and Laplace (recall Theorem 4.1) attract special interest and are studied in the following subsection.

7.1 The normal and the Laplace distribution

The following corollary confirms the vanishing transfer entropy for the case of the multivariate spherically contoured normal distribution.

Corollary 7.1. *The transfer entropy of $W \to U|V$, with $U \oplus V \oplus W \sim \mathcal{N}^d(\mu,\sigma^2\mathbb{I}_d)$, is vanishing, i.e. $TE_{W\to U|V} = 0$.*

Proof. Recall Theorem 4.1 where for value $\gamma := 2$ the $\mathcal{N}_2^d(\mu,\sigma^2\mathbb{I}_d)$ coincides with the d-variate normal $\mathcal{N}_2^d(\mu,\sigma^2\mathbb{I}_d)$. Thus, substituting $g = \gamma/(\gamma-1) := 2$ into (6.3) we obtain that

$$\begin{aligned} TE_{W\to U|V} &= -\frac{1}{2(2\pi)^{d/2}} \sum_{k=0}^{1} \frac{(1)_k}{k!} \sum_{l=0}^{k} \binom{k}{l} (-1)^l 2^{\frac{(k-l)+d}{2}} \Gamma\left(\tfrac{2k-2l+d}{2}\right) \sum_{T_d^{k-l}} \frac{(k-l)!}{t_1!\cdots t_d!} P_d \\ &= -\frac{1}{2(2\pi)^{d/2}} \frac{(1)_0}{0!} \binom{0}{0} 2^{d/2} \Gamma(d/2) \sum_{T_d^0} \frac{1}{t_1!t_2!\cdots t_d!} P_d - \\ &\quad \frac{1}{2(2\pi)^{d/2}} \frac{(1)_1}{1!} \binom{1}{0} 2^{1+d/2} \Gamma(1+d/2) \sum_{T_d^1} \frac{1}{t_1!t_2!\cdots t_d!} P_d - \\ &\quad \frac{1}{2(2\pi)^{d/2}} \frac{(1)_1}{1!} \binom{1}{1} 2^{d/2} \Gamma(d/2) \sum_{T_d^0} \frac{1}{t_1!t_2!\cdots t_d!} P_d, \end{aligned} \tag{7.1}$$

which is true, as the Pochhammer symbol $(1)_k = 0$, for $k > 1$, $k \in \mathbb{N}$, while $(r)_0 := 1$, $r \in \mathbb{R}$. Therefore, (7.1) yields $TE_{W\to U|V} = 0$ as the sets of indexes $T_d^0, T_d^1 = \emptyset$, due to the assumed inequalities $0 < t_1 < k-l = 0$

or $0 < t_{m+2} + t_{m+3} + \cdots + t_d < k - l = 0$ for T_q^0, and $0 < t_1 < k - l = 1$ or $0 < t_{m+2} + t_{m+3} + \cdots + t_d < k - l = 1$for T_d^1 (recall the definition of the indexing set T_d^{k-1}, $d, k - l \in \mathbb{N}$).

□

The following corollary calculates the transfer entropy for the multivariate spherically contoured Laplace distribution.

Corollary 7.2. *The transfer entropy of $W \to U|V$, with $U \oplus V \oplus W \sim \mathscr{L}^{1+m+n}\left(\mu, \sigma^2 \mathbb{I}_{1+m+n}\right)$ is given by*

$$TE_{W \to U|V} = \log \frac{m\,\Gamma\left(\frac{1+m+n}{2}\right)\Gamma\left(\frac{m}{2}\right)}{(m-n)\,\Gamma\left(\frac{1+m}{2}\right)\Gamma\left(\frac{m+n}{2}\right)} - \frac{\Gamma\left(\frac{1+m+n}{2}\right)}{\pi^{(1+m+n)/2}(m+n)!} \times$$
$$\sum_{k=0}^{\infty} \frac{(1/2)_k}{k!} \sum_{l=0}^{k} \binom{k}{l} (-1)^l [2(k-l)+m+n]! \sum_{T_d^{k-l}} \frac{(k-l)!}{t_1! t_2! \cdots t_{1+m+n}!} P_{1+m+n}, \tag{7.2}$$

where P_{1+m+n} is given in (6.4) *when $m, n \in \mathbb{N}$ are both odd or both even numbers, or in* (6.5) *when only one of $m, n \in \mathbb{N}$ is an odd number.*

Proof. Recall Theorem 4.1 for limiting order value $\gamma = \pm\infty$. Then, it holds that the limiting $\mathscr{L}_{\pm\infty}^{1+m+n}(\mu, \Sigma) = \lim_{\gamma \to \pm\infty} \mathscr{N}_\gamma^{1+m+n}(\mu, \Sigma)$, with $\Sigma := \sigma^2 \mathbb{I}_{1+m+n}$, coincides with the d-variate Laplace distribution $\mathscr{L}^d(\mu, \Sigma)$. Thus, substituting $g = \gamma/(\gamma - 1) := 1$ into (6.3) we easily obtain the requested form of (7.2). □

7.2 Kullback-Leibler divergence and the γ-GND

Recall formula (3.7) from Section 3.1. We can express transfer entropy as

$$TE_{Y \to X} = D_{\mathrm{KL}}(p \| q), \tag{7.3}$$

where $p(z) := p(u, v, w)$ with $q(z) := q(u, v, w)$ as in (3.5). Denote by $N_\gamma^d\left(\mu_0, \sigma_1^2 \mathbb{I}_d\right)$ a spherically contoured γ-GND distribution with mean μ_0 and scale matrix $\sigma_1^2 \mathbb{I}_d$. This allows the following conclusions.

Corollary 7.3. *If the joint p.d.f. $p_{(\hat{X}|X^-, Y^-)}(\hat{x}|x^-, y^-)$ exists for stationary random processes Y and X, and for the partial densities holds $p = p(\hat{x}, x^-, y^-) \sim N_\gamma^d\left(\mu_0, \sigma_1^2 \mathbb{I}_d\right)$ and $q = p(\hat{x}, y^-) p(\hat{x}, x^-) [p(x^-)]^{-1} \sim N_\gamma^d\left(\mu_0, \sigma_0^2 \mathbb{I}_d\right)$, then $TE_{Y \to X}$ coincides with KL divergence and can be expressed in terms of parameters of the distribution p and q as*

$$TE_{Y \to X} = D_{\mathrm{KL}}(p \| q) = d \left\{ \log\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\gamma-1}{\gamma} \left[1 - \left(\frac{\sigma_1}{\sigma_0}\right)^{\frac{\gamma}{\gamma-1}} \right] \right\}. \tag{7.4}$$

See [9] for details.

From the above corollary recall that for $\sigma_0 \neq \sigma_1$ the same γ parameter, it is easy to see that

$$D_{\mathrm{KL}}^d(\gamma) < D_{\mathrm{KL}}^{d+1}(\gamma), \quad d = 1, 2, \ldots,$$

where $D_{\mathrm{KL}}^d(\gamma) := D_{\mathrm{KL}}(p \| q)$, $p \sim \mathscr{N}_\gamma^d\left(\mu, \sigma_0^2 \mathbb{I}_d\right)$ and $q \sim \mathscr{N}_\gamma^d\left(\mu, \sigma_1^2 \mathbb{I}_d\right)$. In practice it means, the more variables are involved in processes X and Y, the larger is their KL divergence.

Similar inequalities hold also in case of Laplace probability function for $\gamma \to +\infty$.

For given d, $\mu_1 = \mu_0$, $\sigma_0 \neq \sigma_1$ the KL divergence appears in a strict descending order as $\gamma \in \mathbb{R} \setminus [0\,1]$ rises. In particular,

$$D_{\mathrm{KL}}^d(\gamma_1) > D_{\mathrm{KL}}^d(\gamma_2), \text{ for } \gamma_1 < \gamma_2.$$

Therefore with Laplace, $\gamma \to +\infty$, we obtain a lower bound, i.e.

$$D^d_{\mathrm{KL}}(\infty) < D^d_{\mathrm{KL}}(\gamma), \text{ for every } \gamma \text{ and } d$$

. Moreover, it is easy to see that $TE_{Y\to X}$ is a function of the ratio σ_0/σ_1.

Recall now the definition of the symmetric KL divergence or Jeffreys divergence, [30]:

$$J(p\|q) := \tfrac{1}{2}\left[D_{\mathrm{KL}}(p\|q) + D_{\mathrm{KL}}(q\|p)\right] = \tfrac{1}{2}\int (p-q)(\log p - \log q).$$

Denote with $STE(X,Y)$ the average of $TE_{Y\to X}$ and $TE_{X\to Y}$ for given processes X, Y.

From Corollary 7.3 the following result can be obtained.

Corollary 7.4. *The symmetric transfer entropy $STE(X,Y)$ with $p \sim \mathcal{N}^d_\gamma(\mu, \sigma_0^2\mathbb{I}_d)$ and $q \sim \mathcal{N}^d_\gamma(\mu, \sigma_1^2\mathbb{I}_d)$, equals*

$$STE(X,Y) := \tfrac{1}{2}\left(TE_{Y\to X} + TE_{X\to Y}\right) = -d\,\tfrac{\gamma-1}{2\gamma}\left(2 - s^g - s^{-g}\right), \tag{7.5}$$

where $s := \sigma_1/\sigma_0$ and $g := \gamma/(\gamma-1)$. For p and q normal, it is

$$STE(X,Y) = \tfrac{d}{2}\left(2 - s^2 - s^{-2}\right). \tag{7.6}$$

Proof. Following [30] and evaluating $TE_{Y\to X}$ and $TE_{X\to Y}$ through (7.4), their average gives (7.5) Moreover, the case $\gamma = 2$, again provides the normal case as in (7.6).

□

Notice that when q being a uniform distribution, it holds $STE(X,Y) = 0$, i.e. $TE_{X\to Y}$ and $TE_{Y\to X}$ are cancelling each other. Results can also be obtained for the Geometric and Harmonic means of $D_{\mathrm{KL}}(p\|q)$ and $D_{\mathrm{KL}}(q\|p)$, which provide the corresponding geometric and harmonic means of $TE_{Y\to X}$ and $TE_{X\to Y}$ respectively. These can be useful indexes for their applications. Moreover, as $J(p,q)$ is between min and max of the $D_{\mathrm{KL}}(p\|q)$ and $D_{\mathrm{KL}}(q\|p)$, it provides a measure of comparison about the desired min and max of the corresponding TE.

Corollary 7.5. *It holds that $TE_{Y\times Y_1\to X\times X_1} = TE_{Y\to X} + TE_{Y_1\to X_1}$ with X_1 and Y_1 being also normally distributed random variables.*

Proof. Recall that the Rényi divergence measure R_α is equal to the KL divergence, as Rényi's extra parameter α tends to 1. Then, $R_\alpha(p\times p_1, q\times q_1) = R_\alpha(p,q) + R_\alpha(p_1,q_1)$; see [31]. Therefore, as $\alpha \to 1$, $D_{\mathrm{KL}}(p\times p_1\|q\times q_1) = D_{\mathrm{KL}}(p\|q) + D_{\mathrm{KL}}(p_1\|q_1)$ and hence Corollary is true. □

8 Conclusions

In this paper we evaluated transfer entropy of stationary processes for which we know their probability distributions. Transfer entropy for the normal (Section 5) and for the class of the generalized normal distribution (γ-GND, in Section 6) were computed. Special cases were also discussed in Section 7 and examined for different values of the shape parameter γ of for the γ-GND. We expressed transfer entropy of two processes with their probability distributions given from γ-GND by means of the parameters of these distributions. We derived that the transfer entropy of processes with spherically contoured multivariate normal distributions is zero.

The results for continuous time achieved in the paper can also be applied to the discrete time case, particularly to the time series whose underlying process distribution is from the discussed classes.

In terms of information theory, transfer entropy can be considered as a special case of Kullback-Leibler divergence, namely as a “distance” between two probabilities corresponding to the studied processes.

There are some open problems we still have to face; the easiest one is to study processes for which transfer entropy is always non-zero, i.e. to find a subset of functions, within the family of γ-GND, that their members provide non-zero TE. The simplified (discretized) forms can be applied to time series and therefore can be used in practice. In our future work we shall provide examples with calculations.

Acknowledgement

C. P. Kitsos gratefully acknowledges the Department of Applied Statistics, Johannes Kepler University (JKU) for offering him "Exzellenzstipendium des Landes Oberösterreich" and Katharina Sallinger from JKU for providing useful calculations to him.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Schreiber T. Measuring information transfer. Phys. Rev. Lett. 2000;85.
DOI: 10.1103/PhysRevLett.85.461

[2] Paluš M, Komárek V, Hrnčíř Z, Štěrbová K. Synchronization as adjustment of information rates: Detection from bivariate time series. Phys. Rev. E. 2001;63:046211.

[3] Stramaglia S, Wu G, Pellicoro M, Marinazzo D. Expanding the transfer entropy to identify information circuits in complex systems. Physical Review E. 2012;86:066211.

[4] Hlinka J, Hartman D, Vejmelka M, Runge J, Warwan N, Kurts J, Paluš M. Reliability of inference of directed climate networks using conditional mutual information. Entropy. 2013;15:2023-2045.

[5] Barnett L, Barrett AB, Seth AK. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys. Rev. Letters. 2009;103(23).
DOI: 10.1103/PhysRevLett.103.238701

[6] Prokopenko M, Barnett L, Harré M, Lizier JT, Obst O, Wang XR. Fisher transfer entropy: Quantifying the gain in transient sensitivity. Phil. Trans. R. Soc. A; 2015 (accepted for publication).

[7] Barnett L, Bossomaier T. Transfer entropy as a log-likelihood ratio. Phys. Rev. Letters. 2012;109(13).
DOI: 10.3390/e15062023

[8] Kitsos CP, Tavoularis NK. Logarithmic Sobolev inequalities for information measures. IEEE Transactions on Information Theory. 2009;55(6):2554-2561.

[9] Kitsos CP, Toulias TL. New information measures for the generalized normal distribution. Information. 2010;1:13-27.

[10] Kitsos CP, Toulias TL, Trandafir PC. On the multivariate γ–order normal distribution. Far East Journal of Theoretical Statistics. 2012;38(1):49-73.

[11] Kitsos CP, Vasiliadis VG, Toulias TL. MLE for the γ–order generalized normal distribution. Discussiones Mathematicae - Prob. and Stat. 2014;34:143-158.

[12] Toulias TL, Kitsos CP. Generalizations of entropy and information measures. In: Computation, Cryptography and Network Security, Darras, Rassias (Eds.), Springer. 2015;495-526.

[13] Hlaváčková-Schindler K. Equivalence of Granger causality and transfer entropy: A generalization. Applied Math. Sciences. 2011;5(73):3637-3648.

[14] Tukey J. The Future of data analysis. Ann. Math. Stat. 1961;33(1):1-67.

[15] Pearl J. Causality: Models, reasoning and inference. Economet. Theory, Cambridge Univ. Press; 2003.

[16] Granger CWJ. Investigating causal relations by econometric and crossspectral methods. Econometrica. 1969;37:424-438.

[17] White H, Chalak K, Lu X. Linking Granger causality and the pearl causal model with settable systems. JMLR: Workshop and Conference Proceedings. 2011;12:1-29.

[18] Granger CWJ. Time series analysis, cointegration, and applications. Nobel Lecture; 2003.

[19] Hlaváčková-Schindler K. Gaussianity and non-Gaussianity in natural and social sciences: The assumptions and their influence on causal relationships and their detection in time series. In: Theoretical and Methodological Approaches to Social Sciences and Knowledge Management, Asuncion Lopez-Varela (ed.), Intech. 2012;77-94.

[20] Runge J, Heitzig J, Petoukhov V, Kurths J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. Physical Review Letters. 2012;108.
DOI: 10.1103/PhysRevLett.108.258701

[21] Batu T, Dasgupta S, Kumar R, Rubinfeld R. The complexity of approximating the entropy. In: 34th ACM Symposium on Theory of Computing. 2005;678-687.

[22] Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J. Causality detection based on information-theoretic approaches in time series analysis. Physics Reports. 2007;441:1-46.

[23] Vejmelka M, Hlaváčková-Schindler K. Mutual information estimation in higher dimensions: A speed-up of a k–nearest neighbor based estimator. In: ICANNGA 2007, Part I, Lecture Notes in Computer Science, 4431, Beliczynski et al. (Eds.), Springer Verlag, Berlin Heidelberg. 2007;790-797.

[24] Zhu J, Bellanger JJ, Shu H, Le Bouquin Jeannes R. Contribution to transfer entropy estimation via the k-nearest-neighbors approach. Entropy. 2015;17:4173-4201.

[25] Kaiser A, Schreiber T. Information transfer in continuous processes. Physica D. 2002;166:43-62.

[26] Kullback S, Leibler RA. On information and sufficiency. Annals of Mathematical Statistics. 1951;22(1):79-86.

[27] Cover TM, Thomas JA. Elements of information theory. John Wiley and Sons; 1991.

[28] Toulias TL, Kitsos CP. On the properties of the generalized normal distributions. Discussiones Mathematique - Prob. and Stat. 2012;34:35-49.

[29] Toulias TL. Generalized information for the γ–order normal distribution. Journal of Probability and Statistics. 2015;385285.

[30] Jeffreys H. An invariant form for the prior probability in estimation problems. Proc. Roy. Soc A. 1946;186:453-461.

[31] Harremoes P. Interpretations of Rényi entropies and divergences. Physica A. 2006;365:57-62.

[32] Gradshteyn IS, Ryzhik IM. Table of integrals, series, and products. Elsevier; 2007.

Appendix

A

Proof of Lemma 6.1. Considering the γ-GND r.v.-s $X := U \oplus V \oplus W$, $V := X_v$, $U \oplus V := X_{u\oplus v}$ and $V \oplus W := X_{v\oplus w}$, the corresponding transfer entropy (5.2) gets the form

$$\begin{aligned}TE_{W\to U|V} &= \log \frac{C^d C^m}{C^{1+m} C^{m+1}} - \\ &\frac{C^d}{g} \int\limits_{\mathbb{R}^d} \left(\left\| \frac{\tilde{x}-\mu}{\sigma} \right\|^g + \left\| \frac{x_v-\mu_v}{\sigma} \right\|^g - \left\| \frac{x_{u\oplus v}-\mu_{u\oplus v}}{\sigma} \right\|^g - \left\| \frac{x_{v\oplus w}-\mu_{v\oplus w}}{\sigma} \right\|^g \right) \exp\left\{ -\tfrac{1}{g} \left\| \frac{x-\mu}{\sigma} \right\|^g \right\} \mathrm{d}x,\end{aligned}$$

where $g = g(\gamma) := \gamma/(\gamma-1) > 1$, for all defined γ values, and through the linear transformation $z = z(x) := (x-\mu)/\sigma$, $x \in \mathbb{R}^d$, which implies that $\mathrm{d}x = \sigma^d\,\mathrm{d}z$, we have

$$TE_{W\to U|V} = \log \frac{C^d C^m}{C^{1+m} C^{m+1}} - g^{-1} C^d \sigma^d \int\limits_{\mathbb{R}^d} h_g(z) \exp\left\{ -\tfrac{1}{g} \|z\|^g \right\} \mathrm{d}z, \tag{A.1}$$

where

$$h_g(z) := \|z\|^g + \|z_v\|^g - |z_{u\oplus v}\|^g - \|z_{v\oplus w}\|^g, \quad z \in \mathbb{R}^d, \quad g \in \mathbb{R}_+, \tag{A.2}$$

while $z = (z_i)_{i=1}^p \in \mathbb{R}^d$, $z_v := (z_i)_{i=2}^{m+1} \in \mathbb{R}^m$, $z_{u\oplus v} := z_1 \oplus z_v = (z_i)_{i=1}^{m+1} \in \mathbb{R}^{1+m}$, and $z_{v\oplus w} := z_v \oplus z_w = (z_i)_{i=2}^{1+m+n} \in \mathbb{R}^{m+n}$.

In order to calculate the multiple integral in (A.1), a series expansion of h_g is utilized. Firstly, the series expansion of $f(x) := x^g$, $x, g \in \mathbb{R}_+$ is obtained, i.e.

$$\|z\|^{2g} = f\left(\|z\|^2\right) = \sum_{k=0}^{\infty} \frac{f^{(k)}(1)}{k!} \left(\|z\|^2 - 1\right)^k = \sum_{k=0}^{\infty} \frac{(g)_k}{k!} \left(\|z\|^2 - 1\right)^k, \tag{A.3}$$

$z \in \mathbb{R}^d$, where $(r)_k := r(r-1)(r-2)\cdots(r-k+1)$, $r \in \mathbb{R}$, $k \in \mathbb{N}$, denotes the Pochhammer symbol. Therefore, function h can also be expanded, through (A.3), as

$$h_g(z) = \sum_{k=0}^{\infty} \frac{(g/2)_k}{k!} \left[\left(\|z\|^2 - 1\right)^k + \left(\|z_v\|^2 - 1\right)^k - \left(\|z_{u\oplus v}\|^2 - 1\right)^k - \left(\|z_{v\oplus w}\|^2 - 1\right)^k \right], \tag{A.4}$$

$z \in \mathbb{R}^d$. With the help of the known binomial theorem, the expressions $\left(\|a\|^2 - 1\right)^k$, $a \in \{z, z_v, z_{u\oplus v}, z_{v\oplus w}\}$, $k \in \mathbb{N}$, can also be expanded as

$$\left(\|a\|^2 - 1\right)^k = \sum_{l=0}^{k} \binom{k}{l} (-1)^l \|a\|^{2(k-l)},$$

and hence (A.4) is then written as

$$h_g(z) = \sum_{k=0}^{\infty} \frac{(g/2)_k}{k!} \sum_{l=0}^{k} \binom{k}{l} (-1)^l h_{2(k-l)}(z), \quad z \in \mathbb{R}^d. \tag{A.5}$$

Applying this time the multinomial theorem, the expressions $\|a\|^{2t}$, $a \in \{z, z_v, z_{u,v}, z_{v,w}\}$, can be expanded furthermore as

$$\|z\|^{2t} = \left(\sum_{i=1}^{d} z_i^2\right)^t = \sum_{t_1+t_2+\cdots+t_d=t} A(t_1,t_2,\ldots,t_d), \tag{A.6a}$$

$$\|z_v\|^{2t} = \left(\sum_{i=2}^{m+1} z_i^2\right)^t = \sum_{t_2+t_3+\cdots+t_{m+1}=t} A(t_2,t_3,\ldots,t_{m+1}), \tag{A.6b}$$

$$\|z_{u\oplus v}\|^{2t} = \left(\sum_{i=1}^{m+1} z_i^2\right)^t = \sum_{t_1+t_2+\cdots+t_{m+1}=t} A(t_1,t_2,\ldots,t_{m+1}), \tag{A.6c}$$

$$\|z_{v,\oplus w}\|^{2t} = \left(\sum_{i=2}^{q} z_i^2\right)^t = \sum_{t_2+t_3+\cdots+t_d=t} A(t_2,t_3,\ldots,t_d), \tag{A.6d}$$

where

$$A(t_i)_{i=a}^{b} = A(t_a,t_{a+1},\ldots,t_b) := \frac{t!}{t_a!t_{a+1}!\cdots t_b!} z_a^{2t_a} z_{a+1}^{2t_{a+1}} \cdots z_b^{2t_b}, \tag{A.7}$$

for $a,b = 1,2,\ldots,t$. However, (A.6a), (A.6c), and (A.6d) can be splitted as

$$\begin{aligned}
\|z\|^{2t} &= \sum_{\substack{t_1=t \\ (t_2=\cdots=t_d=0)}} A(t_1) + \sum_{\substack{t_2+t_3+\cdots+t_d=t \\ (t_1=0)}} A(t_2,t_3,\ldots,t_d) + \sum_{\substack{t_1+t_2+\cdots+t_d=t \\ 0<t_1<t \\ 0<t_2+\cdots+t_d<t}} A(t_1,t_2,\ldots,t_d) \\
&= z_1^{2t} + \left[\|z_v\|^{2t} + \sum_{\substack{t_{m+2}+t_{m+3}\cdots+t_d=t \\ (t_1=t_2=\cdots=t_{m+1}=0)}} A(t_{m+2},t_{m+3}\ldots,t_d) + \sum_{\substack{t_2+t_3+\cdots+t_d=t \\ 0<t_2+\cdots+t_{m+1}<t \\ 0<t_{m+2}+\cdots+t_d<t}} A(t_2,t_3,\ldots,t_d) \right] + \sum_{\substack{t_1+t_2+\cdots+t_d=t \\ 0<t_1<t \\ 0<t_2+\cdots+t_d<t}} A(t_1,t_2,\ldots,t_d),
\end{aligned} \tag{A.8a}$$

$$\|z_{u\oplus v}\|^{2t} = z_1^{2t} + \|z_v\|^{2t} + \sum_{\substack{t_1+t_2+\cdots+t_{m+1}=t \\ 0<t_1<t \\ 0<t_2+\cdots+t_{m+1}<t}} A(t_1,t_2,\ldots,t_{m+1}), \tag{A.8b}$$

$$\|z_{v\oplus w}\|^{2t} = \|z_v\|^{2t} + \sum_{\substack{t_{m+2}+\cdots+t_d=t \\ (t_1=\cdots=t_{m+1}=0)}} A(t_{m+2},t_{m+3},\ldots,t_d) + \sum_{\substack{t_2+t_3+\cdots+t_d=t \\ 0<t_2+t_3+\cdots+t_{m+1}<t \\ 0<t_{m+2}+\cdots+t_d<t}} A(t_2,t_3,\ldots,t_d), \tag{A.8c}$$

respectively. We note here that the multi-index inequalities in the above summations has to be considered as "union" rather than as "intersection". For example,

$$\sum_{\substack{t_1+t_2+\cdots+t_d=t \\ \text{(ineq. 1)} \\ \text{(ineq. 2)}}} f(t_1,t_2,\ldots,t_d) := \sum_{\substack{t_1+t_2+\cdots+t_d=t \\ \text{(ineq. 1)}}} f(t_1,t_2,\ldots,t_d) + \sum_{\substack{t_1+t_2+\cdots+t_d=t \\ \text{(ineq. 2)}}} f(t_1,t_2,\ldots,t_d), \tag{A.9}$$

for an arbitrary expression f of d indices $t_1,t_2,\ldots,t_d \in \mathbb{N}$. By substitution of (A.8a)-(A.8c) into (A.2) we derive that

$$
\begin{aligned}
h_{2t}(z) &= \sum_{\substack{t_1+t_2+\cdots+t_d=t\\ 0<t_1<t\\ 0<t_2+t_3+\cdots+t_d<t}} A(t_1,t_2,\dots,t_d) - \sum_{\substack{t_1+t_2+\cdots+t_{m+1}=t\\ 0<t_1<t\\ 0<t_2+t_3+\cdots+t_{m+1}<t}} A(t_1,t_2,\dots,t_{m+1}) \\
&= \left[\sum_{\substack{t_1+t_2+\cdots+t_{m+1}=t\\ 0<t_1<t\\ 0<t_2+t_3+\cdots+t_{m+1}<t\\ (t_{m+2}=\cdots=t_d=0)}} A(t_1,t_2,\dots,t_{m+1}) + \sum_{\substack{t_1+t_2+\cdots+t_d=t\\ 0<t_1<t\\ 0<t_2+t_3+\cdots+t_d<t\\ 0<t_{m+2}+\cdots+t_d<t}} A(t_1,t_2,\dots,t_d)\right] - \sum_{\substack{t_1+t_2+\cdots+t_{m+1}=t\\ 0<t_1<t\\ 0<t_2+t_3+\cdots+t_{m+1}<t}} A(t_1,t_2,\dots,t_{m+1}) = \sum_{\substack{t_1+t_2+\cdots+t_d=t\\ 0<t_1<t\\ 0<t_2+t_3+\cdots+t_d<t\\ 0<t_{m+2}+\cdots+t_d<t}} A(t_1,t_2,\dots,t_d),
\end{aligned}
$$

with $z \in \mathbb{R}^d$, and hence, through (A.7),

$$h_{2t}(z) = \sum_{\substack{t_1+t_2+\cdots+t_d=t\\ 0<t_1<t\\ 0<t_{m+2}+\cdots+t_d<t}} \frac{t!}{t_1!t_2!\cdots t_d!} z_1^{2t_1} z_2^{2t_2} \cdots z_d^{2t_d}, \quad z \in \mathbb{R}^d. \tag{A.10}$$

Therefore, the series expansion of $h_g(z)$ is obtained by substitution of (A.10) to (A.5), i.e.

$$h_g(z) = \sum_{k=0}^{\infty} \frac{(g/2)_k}{k!} \sum_{l=0}^{k} \binom{k}{l} (-1)^l \sum_{\substack{t_1+t_2+\cdots+t_d=k-l\\ 0<t_1<k-l\\ 0<t_{m+2}+\cdots+t_q<k-l}} \frac{(k-l)!}{t_1!t_2!\cdots t_d!} z_1^{2t_1} z_2^{2t_2} \cdots z_d^{2t_d}, \quad z \in \mathbb{R}^d. \tag{A.11}$$

Eventually, the requested transfer entropy as in (A.1) adopts the form

$$TE_{W\to U|V} = \log C - \frac{C^q \sigma^q}{g} \int_{\mathbb{R}^q} \mathrm{e}^{-\frac{1}{g}\|z\|^g} \sum_{k=0}^{\infty} \frac{(g/2)_k}{k!} \sum_{l=0}^{k} \binom{k}{l} (-1)^l \sum_{T_q^{k-l}} \frac{(k-l)!}{t_1!t_2!\cdots t_q!} z_1^{2t_1} z_2^{2t_2} \cdots z_d^{2t_d} \,\mathrm{d}z, \tag{A.12}$$

where $C := C^d C^m / \left(C^{1+m} C^{m+n}\right)$ and T_d^{k-l} is considered to be the set of indices, according to the description in (A.9), as stated in Lemma 6.1.

By switching subsequently to hyperspherical coordinates the multiple integral in (A.12) can be solved, and hence the transfer entropy can finally derived through series expansions. Recall that the known hyperspherical transformation $H : \mathbb{S}_d \to \mathbb{R}^d$, where $\mathbb{S}_d := \mathbb{R}_+ \times [0,\pi)^{d-2} \times [0,2\pi)$ such that

$$\mathbb{S}_d \ni (\rho, \varphi_1, \varphi_2, \dots, \varphi_{d-1}) \xrightarrow{H} (z_1, z_2, \dots, z_d) \in \mathbb{R}^d,$$

is given by

$$
\begin{aligned}
z_1 &= \rho \cos\varphi_1, &\text{(A.13a)}\\
z_i &= \rho \sin\varphi_1 \sin\varphi_2 \cdots \sin\varphi_{i-1} \cos\varphi_i, \quad i = 2,3,\dots,d-1, &\text{(A.13b)}\\
z_q &= \rho \sin\varphi_1 \sin\varphi_2 \cdots \sin\varphi_{d-2} \sin\varphi_{d-1}, &\text{(A.13c)}
\end{aligned}
$$

where $\rho \in \mathbb{R}_+$, $\varphi_1, \varphi_2, \dots, \varphi_{d-2} \in [0,\pi)$, and $\varphi_{d-1} \in [0,2\pi)$. It is hold that $\|z\|^2 = z_1^2 + z_2^2 + \cdots + z_d^2 = \rho^2$, $z \in \mathbb{R}^d$, while the volume element $\mathrm{d}z = \mathrm{d}z_1 \,\mathrm{d}z_2 \cdots \mathrm{d}z_d$ of the d-dimensional Euclidean space is given in hyperspherical coordinates as

$$\mathrm{d}z = \mathrm{J}(H)\,\mathrm{d}\rho\,\mathrm{d}\varphi_1, \varphi_2, \cdots \mathrm{d}\varphi_{d-1} = \rho^{q-1} \left(\prod_{k=1}^{d-2} \sin^{d-k-1}\varphi_k\right) \mathrm{d}\rho\,\mathrm{d}\varphi_1 \cdots \mathrm{d}\varphi_{d-1}, \tag{A.14}$$

where $\mathrm{J}(H)$ is the Jacobian determinant of the transformation H, i.e.

$$\mathrm{J}(H) := \left|\det \frac{\partial(z_1, z_2, \dots, z_d)}{\partial(\rho, \varphi_1, \dots, \varphi_{d-1})}\right| = \rho^{d-1} \sin^{d-2}\varphi_1 \sin^{d-3}\varphi_2 \cdots \sin\varphi_{d-2}. \tag{A.15}$$

Firstly, from (A.13a)-(A.13c) it holds that

$$\begin{aligned} z_1^{2t_1} z_2^{2t_2} \cdots z_d^{2t_d} &= \rho^{2(k-l)} \left[\sin^{2(t_2+\cdots+t_q)} \varphi_1 \cos^{2t_1} \varphi_1\right] \left[\sin^{2(t_3+\cdots+t_d)} \varphi_2 \cos^{2t_2} \varphi_2\right] \cdots \\ &\quad \left[\sin^{2(t_{q-1}+t_d)} \varphi_{q-2} \cos^{2t_{d-2}} \varphi_{d-2}\right] \left[\sin^{2t_d} \varphi_{d-1} \cos^{2t_{d-1}} \varphi_{d-1}\right] \\ &= \rho^{2(k-l)} \prod_{\ell=1}^{d-1} \sin^{2(t_{\ell+1}+\cdots+t_d)} \varphi_\ell \cos^{2t_\ell} \varphi_\ell, \quad z \in \mathbb{R}^d, \end{aligned}$$

with $t_i \in \mathbb{N}_1^d$, $i = 1,2,\ldots,d$, such that $t_1 + t_2 + \cdots + t_d = k - l$, and multiplying the hyperspherically transformed volume element in (A.14) to the above product, we obtain

$$z_1^{2t_1} z_2^{2t_2} \cdots z_d^{2t_d}\, \mathrm{d}z = \rho^{2(k-l)+d-1} \prod_{\ell=1}^{d-1} \sin^{\tau_\ell} \varphi_\ell \cos^{2t_\ell} \varphi_\ell, \quad z \in \mathbb{R}^d, \tag{A.16}$$

where $\tau_\ell := 2(t_{\ell+1} + t_{\ell+2} + \cdots + t_d) + d - \ell - 1$, $\ell = 1,2,\ldots,d-1$.

Applying finally then the hyperspherical transformation (A.13a)-(A.13c) to the multiple integral in (A.12) it is transformed, through (A.16), as in (6.1). □

B

Proof of Theorem 6.2. The transfer entropy from (6.3) can be fully calculated through the calculation of the three multiple integrals (6.2a), (6.2b) and (6.2c).

In particular, the single definite integral in (6.2a) implies, through the appropriate transformation, an expression of the known gamma function, while the definite integrals J_ℓ, $\ell = 1,2,\ldots,d-1$, in (6.2b) and (6.2c), can be calculated using the following indefinite trigonometric integrals, [32, pp. 152–153], i.e.

$$\begin{aligned} \int \sin^\tau \varphi \cos^{2n} \mathrm{d}\varphi &= \frac{\sin^{\tau+1}\varphi}{2n+\tau} \left\{ \cos^{2n-1}\varphi + \sum_{k=1}^{n-1} \frac{(2n-1)(2n-3)\cdots(2n-2k+1)\cos^{2(n-k)-1}\varphi}{(2n+\tau-2)(2n+\tau-4)\cdots(2n+\tau-2k)} \right\} + \\ &\quad \frac{(2n-1)!!}{(2n+\tau)(2n+\tau-2)\cdots(\tau+2)} \int \sin^\tau \varphi\, \mathrm{d}\varphi, \quad n \in \mathbb{N}, \quad \tau \in \mathbb{R} \setminus \{-2,-4,\ldots,-2n\}, \end{aligned} \tag{B.1}$$

and

$$\int \sin^{2n} \varphi\, \mathrm{d}\varphi = \frac{1}{2^{2n}} \binom{2n}{n} \varphi + \frac{(-1)^n}{2^{2n-1}} \sum_{k=0}^{n-1} \binom{2n}{k} \frac{(-1)^k \sin\big(2(n-k)\varphi\big)}{2(n-k)}, \tag{B.2a}$$

$$\int \sin^{2n+1} \varphi\, \mathrm{d}\varphi = \frac{(-1)^{n+1}}{2^{2n}} \sum_{k=0}^{n} \binom{2n+1}{k} \frac{(-1)^k \cos\big(2(n-k)\varphi + \varphi\big)}{2(n-k)+1}, \tag{B.2b}$$

Applying the transformation $r = r(\rho) := g^{-1}\rho^g$, $\rho \in \mathbb{R}_+$, which yields $\mathrm{d}r = \rho^{d-1}\mathrm{d}\rho$, integral (6.2a) is calculated as

$$\begin{aligned} I_{k-l} &= \int_0^{+\infty} \rho^{2(k-l)+d-g} \mathrm{e}^{-r}\, \mathrm{d}r = g^{\frac{2(k-l)+d}{g}-1} \int_0^{+\infty} \left(g^{-1}\rho^g\right)^{\frac{2(k-l)+d-g}{g}} \mathrm{e}^{-r}\, \mathrm{d}r \\ &= g^{\frac{2(k-l)+d}{g}-1} \int_0^{+\infty} r^{-1+\frac{2(k-l)+d}{g}} \mathrm{e}^{-r}\, \mathrm{d}r = g^{\frac{2(k-l)+d}{g}-1} \Gamma\left(\tfrac{2(k-l)+d}{g}\right). \end{aligned} \tag{B.3}$$

As far as the trigonometric integrals, as in (6.2b) and (6.2c), are concerned, we note that due to the fact that the power values τ_ℓ, $\ell = 1,2,\ldots,d-1$, can be either odd or even numbers, they have an influence on the corresponding integration, as (B.1) needs the calculation of $\int \sin^{\tau_\ell} \varphi_\ell\, \mathrm{d}\varphi_\ell$, which is given through (B.2a) or (B.2a). Therefore, the product of J_ℓ in (6.1) has to be splitted in odd and even parts in order the right form of (B.2a) or (B.2a) can be used. In particular, the following product-splits are adopted:

- **Case $d := 2s+1$, $s \in \mathbb{N}$:** Equivalently $m,n \in \mathbb{N}$ being both odd or even numbers. It holds that

$$\prod_{\ell=1}^{d-1:=2s} J_\ell = (J_1J_3\cdots J_{2s-1})(J_2J_4\cdots J_{2s-2})J_{2s} = \left(\prod_{\ell=0}^{\frac{d-1}{2}-1} J_{2\ell+1}\right)\left(\prod_{\ell=1}^{\frac{d-1}{2}-1} J_{2\ell}\right)J_{d-1}, \tag{B.4}$$

while the power values $\tau_{2\ell+1} = 2(t_{2\ell+2}+t_{2\ell+3}+\cdots+t_d)+d-2\ell$, $\ell = 0,1,\ldots,(d-1)/2-1$, are all then odd numbers (recall the definition of t_ℓ, $\ell = 1,2,\ldots,d-1$, in (A.16)). Thus, integrals $J_{2\ell+1}$, $\ell = 0,1,\ldots,(d-1)/2-1$, as in (6.2b), are then calculated through (B.1) and (B.2b), i.e.

$$\begin{aligned} J_{2\ell+1} &= \left[\int \sin^{\tau_{2\ell+1}}\varphi_{2\ell+1}\cos^{2t_{2\ell+1}}\varphi_{2\ell+1}\,\mathrm{d}\varphi_{2\ell+1}\right]_{\varphi_{2\ell+1}=0}^{\pi} \\ &= 0+\frac{(2t_{2\ell+1}-1)!!\left[\int \sin^{\tau_{2\ell+1}}\varphi_{2\ell+1}\,\mathrm{d}\varphi_{2\ell+1}\right]_{\varphi_{2\ell+1}=0}^{\pi}}{(2t_{2\ell+1}+\tau_{2\ell+1})(2t_{2\ell+1}+\tau_{2\ell+1}-2)\cdots(\tau_{2\ell+1}+2)} \\ &= \frac{(-1)^{(\tau_{2\ell+1}+1)/2}(2t_{2\ell+1}-1)!!}{2^{\tau_{2\ell+1}-2}(2t_{2\ell+1}+\tau_{2\ell+1})(2t_{2\ell+1}+\tau_{2\ell+1}-2)\cdots(\tau_{2\ell+1}+2)}\sum_{j=0}^{(\tau_{2\ell-1}-1)/2}\frac{(-1)^j}{\tau_{2\ell+1}-2j}\binom{\tau_{2\ell+1}}{j}, \end{aligned} \tag{B.5}$$

for $\ell = 0,1,\ldots,(d-1)/2-1$. Moreover, as $\tau_{2\ell}$, $\ell = 1,2,\ldots,(d-1)/2-1$, and τ_{d-1} are even number, integrals J_{d-1} as in (6.2b), and $J_{2\ell}$, $\ell = 1,2,\ldots,(d-1)/2-1$, as in (6.2c), are then calculated through (B.1) and (B.2a), i.e.

$$\begin{aligned} J_{d-1} &= \left[\int \sin^{\tau_{d-1}}\varphi_{d-1}\cos^{2t_{d-1}}\varphi_{d-1}\,\mathrm{d}\varphi_{d-1}\right]_{\varphi_{d-1}=0}^{2\pi} = 0+\frac{(2t_{d-1}-1)!!\left[\int \sin^{\tau_{d-1}}\varphi_{d-1}\,\mathrm{d}\varphi_{d-1}\right]_{\varphi_{d-1}=0}^{2\pi}}{(2t_{d-1}+\tau_{d-1})(2t_{d-1}+\tau_{d-1}-2)\cdots(\tau_{d-1}+2)} \\ &= \frac{\pi(2t_{d-1}-1)!!}{2^{\tau_{d-1}-1}(2t_{d-1}+\tau_{d-1})(2t_{d-1}+\tau_{d-1}-2)\cdots(\tau_{d-1}+2)}\binom{\tau_{d-1}}{\tau_{d-1}/2}, \text{ and} \end{aligned} \tag{B.6}$$

$$J_{2\ell} = \left[\int \sin^{\tau_{2\ell}}\varphi_{2\ell}\cos^{2t_{2\ell}}\varphi_{2\ell}\,\mathrm{d}\varphi_{2\ell}\right]_{\varphi_{2\ell}=0}^{\pi} = \frac{\pi(2t_{2\ell}-1)!!}{2^{\tau_{2\ell}}(2t_{2\ell}+\tau_{2\ell})\cdots(\tau_{2\ell}+2)}\binom{\tau_{2\ell}}{\tau_{2\ell}/2}, \tag{B.7}$$

for $\ell = 1,2,\ldots,(d-3)/2$. Thus, by substitution of (B.5)–(B.7) into (B.4), we derive that $\prod_{\ell=1}^{d-1} J_\ell = P_d$, with P_d as in (6.4), where R_k, $k \in \mathbb{N}$, as in (6.6). Note that the last equality in (6.6) holds from the fact that $(r/2)_k = r(r-2)(r-4)\cdots(r-2k+2)/(r^k)$, $r \in \mathbb{R}$, $k \in \mathbb{N}$, and hence the product $(2n+\tau)(2n+\tau-2)\cdots(\tau+2)$ as appeared in (B.1) can be written, in a more compact form, as $\left(\frac{2n+\tau}{2}\right)_n(2n+\tau)^n$.

- **Case $d := 2s$, $s \in \mathbb{N}$:** Equivalently $m,n \in \mathbb{N}$ not being both odd or even numbers simultaneously. It holds that

$$\prod_{\ell=1}^{d-1:=2s-1} J_\ell = (J_1J_3\cdots J_{2s-3})J_{2s-1}(J_2J_4\cdots J_{2s-2}) = \left(\prod_{\ell=0}^{\frac{d}{2}-2} J_{2\ell+1}\right)J_{d-1}\left(\prod_{\ell=1}^{\frac{d}{2}-1} J_{2\ell}\right), \tag{B.8}$$

while the power values $\tau_{2\ell+1} = 2(t_{2\ell+2}+t_{2\ell+3}+\cdots+t_d)+d-2\ell$, $\ell = 0,1,\ldots,d/2-1$, correspond then to even numbers. Thus, integrals $J_{2\ell+1}$, $\ell = 0,1,\ldots,d/2-1$, as in (6.2b), are then calculated through (B.1) and (B.2b), i.e.

$$\begin{aligned} J_{d-1} &= \left[\int \sin^{\tau_{d-1}}\varphi_{d-1}\cos^{2t_{d-1}}\varphi_{d-1}\,\mathrm{d}\varphi_{d-1}\right]_{\varphi_{d-1}=0}^{2\pi} \\ &= \frac{\pi(2t_{d-1}-1)!!}{2^{\tau_{d-1}-1}(2t_{d-1}+\tau_{d-1})(2t_{d-1}+\tau_{d-1}-2)\cdots(\tau_{d-1}+2)}\binom{\tau_{d-1}}{\tau_{d-1}/2}, \text{ and} \end{aligned} \tag{B.9}$$

$$\begin{aligned} J_{2\ell+1} &= \left[\int \sin^{\tau_{2\ell+1}}\varphi_{2\ell+1}\cos^{2t_{2\ell+1}}\varphi_{2\ell+1}\,\mathrm{d}\varphi_{2\ell+1}\right]_{\varphi_{2\ell+1}=0}^{\pi} \\ &= \frac{\pi(2t_{2\ell+1}-1)!!}{2^{\tau_{2\ell+1}}(2t_{2\ell+1}+\tau_{2\ell+1})(2t_{2\ell+1}+\tau_{2\ell+1}-2)\cdots(\tau_{2\ell+1}+2)}\binom{\tau_{2\ell+1}}{\tau_{2\ell+1}/2}, \quad \ell = 0,1,\ldots,\tfrac{d}{2}-2. \end{aligned} \tag{B.10}$$

Moreover, as $\tau_{2\ell}$, $\ell = 1,2,\ldots,d/2-1$, are odd numbers, integrals $J_{2\ell}$, $\ell = 1,2,\ldots,d/2-1$, as in (6.2b), are then calculated through (B.1) and (B.2b), i.e.

$$J_{2\ell} = \left[\int \sin^{\tau_{2\ell}} \varphi_{2\ell} \cos^{2t_{2\ell}} \varphi_{2\ell}\, \mathrm{d}\varphi_{2\ell}\right]_{\varphi_{2\ell}=0}^{\pi} = \frac{(-1)^{(\tau_{2\ell}+1)/2}(2t_{2\ell}-1)!!}{2^{\tau_{2\ell}-2}(2t_{2\ell}+\tau_{2\ell})\cdots(\tau_{2\ell}+2)} \sum_{j=0}^{(\tau_{2\ell-1}-1)/2} \frac{(-1)^j}{\tau_{2\ell}-2j}\binom{\tau_{2\ell}}{j}, \quad \text{(B.11)}$$

for $\ell = 1,2,\ldots,\frac{d}{2}-1$. Thus, by substitution of (B.9)–(B.11) into (B.8), we derive that $\prod_{\ell=1}^{d-1} J_\ell = P_d$, with P_d as in (6.5).

Therefore, by substitution of (6.2a), and $\prod_{\ell=1}^{d-1} J_\ell$ with P_d (for d odd as in (6.4), or for d even as in (6.5)) into (6.1), and then using (4.4), the requested transfer entropy is eventually derived in (6.3) as a series expansion form.

□

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/15385