

Article

Transfer Entropy Expressions for a Class of Non-Gaussian Distributions

Mehrdad Jafari-Mamaghani ^{1,2,*} and Joanna Tyrcha ¹

¹ Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden;
E-Mail: joanna@math.su.se

² Center for Biosciences, Department of Biosciences and Nutrition, Karolinska Institutet, SE-141 83 Huddinge, Sweden

* Author to whom correspondence should be addressed; E-Mail: mjm@math.su.se; Tel.: +46-8-164507.

Received: 17 January 2014; in revised form: 10 March 2014 / Accepted: 18 March 2014 /

Published: 24 March 2014

Abstract: Transfer entropy is a frequently employed measure of conditional co-dependence in non-parametric analysis of Granger causality. In this paper, we derive analytical expressions for transfer entropy for the multivariate exponential, logistic, Pareto (type $\mathcal{I} - \mathcal{IV}$) and Burr distributions. The latter two fall into the class of fat-tailed distributions with power law properties, used frequently in biological, physical and actuarial sciences. We discover that the transfer entropy expressions for all four distributions are identical and depend merely on the multivariate distribution parameter and the number of distribution dimensions. Moreover, we find that in all four cases the transfer entropies are given by the same decreasing function of distribution dimensionality.

Keywords: Granger causality; information theory; transfer entropy; multivariate distributions; power-law distributions

1. Introduction

Granger causality is a well-known concept based on dynamic co-dependence [1]. In the framework of Granger causality, the cause precedes and contains unique information about the effect. The concept of Granger causality has been applied in a wide array of scientific disciplines from econometrics to neurophysiology, from sociology to climate research (see [2,3] and references therein), and most recently in cell biology [4].

Information theory has increasingly become a useful complement to the existing repertoire of methodologies in mathematical statistics [5,6]. Particularly, in the area of Granger causality, transfer entropy [7], an information theoretical measure of co-dependence based on Shannon entropy, has been applied extensively in non-parametric analysis of time-resolved causal relationships. It has been shown that (conditional) mutual information measured in *nats* and transfer entropy coincide in definition [8–10]. Moreover, for Gaussian-distributed variables, there is a tractable equivalence by a factor of two between transfer entropy and a linear test statistic for Granger causality [11]. Although similar equivalences for non-Gaussian variables have been given in [8], it should be remarked that such equivalences cannot be generalized to non-Gaussian distributions as the linear models underlying the construction of linear test statistics for Granger causality are rendered invalid under assumptions of non-Gaussianity.

The aim of this paper is to present closed-form expressions for transfer entropy for a number of non-Gaussian, unimodal, skewed distributions used in the modeling of occurrence rates, rare events and ‘fat-tailed’ phenomena in biological, physical and actuarial sciences [12]. More specifically, we will derive expressions for transfer entropy for the multivariate exponential, logistic, Pareto (type $\mathcal{I} - \mathcal{IV}$) and Burr distributions. As for real-world applications, the exponential distribution is the naturally occurring distribution for describing inter-arrival times in a homogeneous Poisson process. In a similar manner, the exponential distribution can be used to model many other change of state scenarios in continuous settings, e.g., time until the occurrence of an accident given certain specifications. The logistic distribution is of great utility given its morphological similarity to the Gaussian distribution and is frequently used to model Gaussian-like phenomena in the presence of thicker distribution tails. The Pareto distribution (in either of its forms) is used in modeling of size related phenomena such as size of incurred casualties in non-life insurance, size of meteorites, and size of trafficked files over the Internet. The Burr distribution is another distribution used in non-life insurance to model incurred casualties, as well as in econometrics where it is used to model income distribution.

The specific choice of these distributions is contingent upon the existence of unique expressions for the corresponding probability density functions and Shannon entropy expressions. A counter-example is given by the multivariate gamma distribution, which although derived in a number of tractable formats under certain preconditions [12,13], lacks a unique and unequivocal multivariate density function and hence a unique Shannon entropy expression.

Another remark shall be dedicated to stable distributions. Such distributions are limits of appropriately scaled sums of independent and identically distributed variables. The general tractability of distributions with this property lies in their “attractor” behavior and their ability to accommodate skewness and heavy tails. Other than the Gaussian distribution (stable by the Central Limit Theorem), the Cauchy-Lorentz distribution and the Lévy distribution are considered to be the only stable distributions that can be expressed analytically. However, the latter lacks analytical expressions for Shannon entropy in the multivariate case. Expressions for Shannon entropy and transfer entropy for the multivariate Gaussian distribution have been derived in [14] and [11], respectively. Expressions for Shannon entropy and transfer entropy for the multivariate Cauchy-Lorentz distribution can be found in the Appendix.

As a brief methodological introduction, we will go through a conceptual sketch of Granger causality, the formulation of the linear models underlying the above-mentioned test statistic, and the definition of transfer entropy before deriving the expressions for our target distributions.

2. Methods

Employment of Granger causality is common practice within cause-effect analysis of dynamic phenomena where the cause temporally precedes the effect and where the information embedded in the cause about the effect is unique. Formulated using probability theory, under H_0 , given k lags and the random variables A and B and the set of all other random variables C in any arbitrary system, B is said to not Granger-cause A at observation index t , if

$$H_0 : A_t \perp\!\!\!\perp \{B_{t-1}, \dots, B_{t-k}\} | \{A_{t-1}, \dots, A_{t-k}, C_{t-1}, \dots, C_{t-k}\} \quad (1)$$

where $\perp\!\!\!\perp$ denotes probabilistic independence. Henceforth, for the sake of convenience, we implement the following substitutions: $X = A_t$, $Y = \{B\}_{t-1}^{t-k}$ and $Z = \{A, C\}_{t-1}^{t-k}$. It is understood that all formulations in what follows are compatible with any multivariate setting. Thus, one can parsimoniously reformulate the hypothesis in Equation (1) as:

$$H_0 : X \perp\!\!\!\perp Y | Z \quad (2)$$

The statement above can be tested by comparing the two conditional probability densities: $f_{X|Z}$ and $f_{X|YZ}$ [15].

2.1. Linear Test Statistics

In parametric analysis of Granger causality, techniques of linear regression have been the dominant choice. Under fulfilled assumptions of ordinary least squares' regression and stationarity, the hypothesis in Equation (2), can be tested using the following models:

$$H_0 : X = \beta_1 + Z\beta_2 + \epsilon \quad (3)$$

$$H_1 : X = \gamma_1 + Z\gamma_2 + Y\gamma_3 + \eta \quad (4)$$

where the β and γ terms are the regression coefficients, and the residuals ϵ and η are independent and identically distributed following a centered Gaussian $N(0, \sigma^2)$. Traditionally, the F-distributed Granger-Sargent test [1], equivalent to the structural Chow test [16], has been used to examine the statistical significance of the reduction in residual sum of squares in the latter model compared to the former. In this study however, we will focus on the statistic $G(X, Y|Z) = \ln(\text{Var}_\epsilon / \text{Var}_\eta)$ [11,17]. This statistic is χ^2 -distributed under the null hypothesis, and non-central χ^2 -distributed under the alternate hypothesis. There are two types of multivariate generalizations of $G(X, Y|Z)$; one by means of *total variance*, using the trace of covariance matrices [18], and one by *generalized variance*, using the determinant of covariance matrices [11,17]. For a thorough discussion on the advantages of either measure we refer the reader to [18,19]. Choosing the latter extension, the test statistic in $G(X, Y|Z)$ can be reformulated as:

$$\begin{aligned} G(X, Y|Z) &= \ln \left(\frac{|\Sigma_\epsilon|}{|\Sigma_\eta|} \right) \\ &= \ln \left(\frac{|\Sigma_{XZ}| \cdot |\Sigma_{YZ}|}{|\Sigma_Z| \cdot |\Sigma_{XYZ}|} \right) \end{aligned} \quad (5)$$

where the last equality follows the scheme presented in [11].

2.2. Transfer Entropy

Transfer entropy, a non-parametric measure of co-dependence is identical to (conditional) mutual information measured in *nats* (using the natural logarithm). Mutual information is a basic concept, based on the most fundamental measure in information theory, the Shannon entropy, or, more specifically, the differential Shannon entropy in the case of continuous distributions. The differential Shannon entropy of a random variable S with a continuous probability density f_S with support on \mathcal{S} is

$$H(S) \equiv -E[\log_b f_S] = - \int_{\mathcal{S}} f_S \log_b f_S ds \tag{6}$$

where b is the base of the logarithm determining the terms in which the entropy is measured; $b = 2$ for *bits* and $b = e$ for *nats* [14,20]. The transfer entropy for the hypothesis in Equation (2) is defined as [7]:

$$\begin{aligned} T(Y \rightarrow X|Z) &= H(X|Z) - H(X|Y, Z) \\ &= H(X, Z) - H(Z) + H(Y, Z) - H(X, Y, Z) \end{aligned} \tag{7}$$

Interestingly, for Gaussian variables one can show that $G(X, Y|Z) = 2 \cdot T(Y \rightarrow X|Z)$ [11]. Naturally, such equivalences fail when using other types of distributions that do not meet the requirements of linear models used to construct $G(X, Y|Z)$.

In the following, we shall look at closed-form expressions for transfer entropy for the multivariate exponential, logistic, Pareto (type $\mathcal{I} - \mathcal{IV}$) and Burr distributions. Before deriving the results, it should be noted that all marginal densities of the multivariate density functions in this study are distributed according to the same distribution; *i.e.*, the marginal densities of a multivariate exponential density are themselves exponential densities.

3. Results

In this section we will derive the expression for transfer entropy for the multivariate exponential distribution. The remaining derivations follow an identical scheme and are presented in the Appendix. The differential Shannon entropy expressions employed in this study can be found in [21].

The multivariate exponential density function for a d -dimensional random vector \mathbf{S} is:

$$f_{\mathbf{S}} = \prod_{i=1}^d \frac{\alpha + i - 1}{\theta_i} \exp\left(-\frac{s_i - \lambda_i}{\theta_i}\right) \left[\sum_{i=1}^d \exp\left(-\frac{s_i - \lambda_i}{\theta_i}\right) - d + 1 \right]^{-(\alpha+d)} \tag{8}$$

where $\mathbf{S} \in \mathbb{R}^d$, $s_i > \lambda_i$, $\theta_i > 0$ for $i = 1, \dots, d$ and $\alpha > 0$. For the multivariate exponential distribution the differential Shannon entropy of \mathbf{S} is:

$$H(\mathbf{S}) = - \sum_{i=1}^d \log\left(\frac{\alpha + i - 1}{\theta_i}\right) + (\alpha + d) \sum_{i=1}^d \frac{1}{\alpha + i - 1} - \frac{d}{\alpha} \tag{9}$$

Thus, transfer entropy for a set of multivariate exponential variables can be formulated as:

$$\begin{aligned}
 T(Y \rightarrow X|Z) &= H(X, Z) - H(Z) + H(Y, Z) - H(X, Y, Z) \\
 &= - \sum_{i=1}^{d_X} \log \left(\frac{\alpha + i - 1}{\theta_i^{(X)}} \right) - \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + d_X + i - 1}{\theta_i^{(Z)}} \right) \\
 &\quad + (\alpha + d_X + d_Z) \left(\sum_{i=1}^{d_X} \frac{1}{\alpha + i - 1} + \sum_{i=1}^{d_Z} \frac{1}{\alpha + d_X + i - 1} \right) \\
 &\quad - \frac{d_X + d_Z}{\alpha} + \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + i - 1}{\theta_i^{(Z)}} \right) - (\alpha + d_Z) \left(\sum_{i=1}^{d_Z} \frac{1}{\alpha + i - 1} \right) \\
 &\quad + \frac{d_Z}{\alpha} - \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + i - 1}{\theta_i^{(Z)}} \right) - \sum_{i=1}^{d_Y} \log \left(\frac{\alpha + d_Z + i - 1}{\theta_i^{(Y)}} \right) \\
 &\quad + (\alpha + d_Z + d_Y) \left(\sum_{i=1}^{d_Z} \frac{1}{\alpha + i - 1} + \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_Z + i - 1} \right) \\
 &\quad - \frac{d_Z + d_Y}{\alpha} + \sum_{i=1}^{d_X} \log \left(\frac{\alpha + i - 1}{\theta_i^{(X)}} \right) + \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + d_X + i - 1}{\theta_i^{(Z)}} \right) \\
 &\quad + \sum_{i=1}^{d_Y} \log \left(\frac{\alpha + d_X + d_Z + i - 1}{\theta_i^{(Y)}} \right) - (\alpha + d_X + d_Z + d_Y) \\
 &\quad \left(\sum_{i=1}^{d_X} \frac{1}{\alpha + i - 1} + \sum_{i=1}^{d_Z} \frac{1}{\alpha + d_X + i - 1} + \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_X + d_Z + i - 1} \right) \\
 &\quad + \frac{d_X + d_Z + d_Y}{\alpha}
 \end{aligned} \tag{10}$$

which, after simplifications, reduces to

$$\begin{aligned}
 T(Y \rightarrow X|Z) &= \sum_{i=1}^{d_Y} \log \left(1 + \frac{d_X}{\alpha + d_Z + i - 1} \right) \\
 &\quad - d_Y \left[\sum_{i=1}^{d_X} \frac{1}{\alpha + i - 1} + \sum_{i=1}^{d_Z} \left(\frac{1}{\alpha + d_X + i - 1} - \frac{1}{\alpha + i - 1} \right) \right] \\
 &\quad + (\alpha + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_Z + i - 1} \\
 &\quad - (\alpha + d_X + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_X + d_Z + i - 1}
 \end{aligned} \tag{11}$$

where d_X represents the number of dimensions in X , and where α is the multivariate distribution parameter. As stated previously, the expression in Equation (11) holds for the multivariate logistic, Pareto (type $\mathcal{I} - \mathcal{IV}$) and Burr distributions as proven in the Appendix. For the specific case of $d_X = d_Y = d_Z = 1$, the transfer entropy expression reduces to:

$$T(Y \rightarrow X|Z) = \log \left(\frac{\alpha + 2}{\alpha + 1} \right) - \frac{1}{\alpha + 2} \tag{12}$$

In any regard, $T(Y \rightarrow X|Z)$ depends only on the number of involved dimensions and the parameter α . The latter parameter, α , operates as a multivariate distribution feature and does not have a univariate counterpart. This result indicates that the value assigned to the conditional transfer of information from the cause to the effect decreases with increasing values of α . However, the impact of the multivariate distribution parameter α in this decrease, shrinks rather rapidly as the numbers of dimensions increase.

4. Conclusions

The distributions discussed in this paper are frequently subject to the modeling of natural phenomena, and utilized frequently within biological, physical and actuarial engineering. Events distributed according to any of the discussed distributions are not suitable for analysis using linear models and require non-parametric models of analysis or transformations where feasible.

The focus of this paper has been on non-parametric modeling of Granger causality using transfer entropy. Our results show that the expressions for transfer entropy for the multivariate exponential, logistic, Pareto (type $\mathcal{I} - \mathcal{IV}$) and Burr distributions coincide in definition and are dependent on the multivariate distribution parameter α , and the number of dimensions. In other words, the transfer entropy expressions are independent of other parameters of the multivariate distributions.

As underlined by our result, the value of transfer entropy depends in a declining manner on the multivariate distribution parameter α as the number of dimensions increase.

Acknowledgments

The authors wish to thank John Hertz for insightful discussions and feedback. MJM has been supported by the Magnussons Fund at the Royal Swedish Academy of Sciences and the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement #258068, EU-FP7-Systems Microscopy NoE. MJM and JT have been supported by the Swedish Research Council grant #340-2012-6011.

Author Contributions

Mehrdad Jafari-Mamaghani and Joanna Tyrcha designed, performed research and analyzed the data; Mehrdad Jafari-Mamaghani wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

A. Appendix

A.1. Multivariate Logistic Distribution

The multivariate logistic density function for a d -dimensional random vector \mathbf{S} is:

$$f_{\mathbf{S}} = \prod_{i=1}^d \frac{\alpha + i - 1}{\theta_i} \exp\left(-\frac{s_i - \lambda_i}{\theta_i}\right) \left[\sum_{i=1}^d \exp\left(-\frac{s_i - \lambda_i}{\theta_i}\right) + 1 \right]^{-(\alpha+d)} \quad (13)$$

with $\mathbf{S} \in \mathbb{R}^d$, $\theta_i > 0$ for $i = 1, \dots, d$ and $\alpha > 0$. For the multivariate logistic distribution the differential Shannon entropy of \mathbf{S} is:

$$H(\mathbf{S}) = - \sum_{i=1}^d \log\left(\frac{\alpha + i - 1}{\theta_i}\right) + (\alpha + d)\Psi(\alpha + d) - \alpha\Psi(\alpha) - d\Psi(1) \quad (14)$$

where $\Psi(s) = \frac{d}{ds} \ln \Gamma(s)$ is the digamma function. Thus, the transfer entropy for the multivariate logistic distribution can be formulated as:

$$\begin{aligned} T(Y \rightarrow X|Z) &= H(XZ) - H(Z) + H(YZ) - H(XYZ) \\ &= - \sum_{i=1}^{d_X} \log\left(\frac{\alpha + i - 1}{\theta_i^{(X)}}\right) - \sum_{i=1}^{d_Z} \log\left(\frac{\alpha + d_X + i - 1}{\theta_i^{(Z)}}\right) \\ &\quad + (\alpha + d_X + d_Z)\Psi(\alpha + d_X + d_Z) - \alpha\Psi(\alpha) - (d_X + d_Z)\Psi(1) \\ &\quad + \sum_{i=1}^{d_Z} \log\left(\frac{\alpha + i - 1}{\theta_i^{(Z)}}\right) \\ &\quad - (\alpha + d_Z)\Psi(\alpha + d_Z) + \alpha\Psi(\alpha) + d_Z\Psi(1) \\ &\quad - \sum_{i=1}^{d_Z} \log\left(\frac{\alpha + i - 1}{\theta_i^{(Z)}}\right) - \sum_{i=1}^{d_Y} \log\left(\frac{\alpha + d_Z + i - 1}{\theta_i^{(Y)}}\right) \\ &\quad + (\alpha + d_Z + d_Y)\Psi(\alpha + d_Z + d_Y) - \alpha\Psi(\alpha) - (d_Z + d_Y)\Psi(1) \\ &\quad + \sum_{i=1}^{d_X} \log\left(\frac{\alpha + i - 1}{\theta_i^{(X)}}\right) + \sum_{i=1}^{d_Z} \log\left(\frac{\alpha + d_X + i - 1}{\theta_i^{(Z)}}\right) \\ &\quad - \sum_{i=1}^{d_Y} \log\left(\frac{\alpha + d_X + d_Z + i - 1}{\theta_i^{(Y)}}\right) \\ &\quad - (\alpha + d_X + d_Z + d_Y)\Psi(\alpha + d_X + d_Z + d_Y) \\ &\quad + \alpha\Psi(\alpha) + (d_X + d_Z + d_Y)\Psi(1) \end{aligned} \quad (15)$$

which, after simplifications, using the identity

$$\Psi(\alpha + d) = \Psi(\alpha) + \sum_{i=1}^d \frac{1}{\alpha + i - 1} \quad (16)$$

reduces to

$$T(Y \rightarrow X|Z) = \sum_{i=1}^{d_Y} \log\left(1 + \frac{d_X}{\alpha + d_Z + i - 1}\right)$$

$$\begin{aligned}
 & -d_Y \left[\sum_{i=1}^{d_X} \frac{1}{\alpha + i - 1} + \sum_{i=1}^{d_Z} \left(\frac{1}{\alpha + d_X + i - 1} - \frac{1}{\alpha + i - 1} \right) \right] \\
 & + (\alpha + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_Z + i - 1} \\
 & - (\alpha + d_X + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_X + d_Z + i - 1}
 \end{aligned} \tag{17}$$

A.2. Multivariate Pareto Distribution

The multivariate Pareto density function of type $\mathcal{I} - \mathcal{IV}$ for a d -dimensional random vector \mathbf{S} is:

$$f_{\mathbf{S}} = \prod_{i=1}^d \frac{\alpha + i - 1}{\gamma_i \theta_i} \left(\frac{s_i - \mu_i}{\theta_i} \right)^{(1/\gamma_i)-1} \left(1 + \sum_{i=1}^d \left(\frac{s_i - \mu_i}{\theta_i} \right)^{1/\gamma_i} \right)^{-(\alpha+d)} \tag{18}$$

with $\mathbf{S} \in \mathbb{R}^d$, $s_i > \mu_i$, $\gamma_i > 0$ and $\theta_i > 0$ for $i = 1, \dots, d$ and $\alpha > 0$. Other types of the multivariate Pareto density function are obtained as follows:

- Pareto \mathcal{III} by setting $\alpha = 1$ in Equation (18).
- Pareto \mathcal{II} by setting $\gamma_i = 1$ in Equation (18).
- Pareto \mathcal{I} by setting $\gamma_i = 1$ and $\mu_i = \theta_i$ in Equation (18).

For the multivariate Pareto distribution in Equation (18) the differential entropy of \mathbf{S} is:

$$H(\mathbf{S}) = - \sum_{i=1}^d \log \left(\frac{\alpha + i - 1}{\gamma_i \theta_i} \right) + (\alpha + d) [\Psi(\alpha + d) - \Psi(\alpha)] - [\Psi(1) - \Psi(\alpha)] \left(d - \sum_{i=1}^d \gamma_i \right) \tag{19}$$

Thus, the transfer entropy for the multivariate Pareto density function of type $\mathcal{I} - \mathcal{IV}$ can be formulated as:

$$\begin{aligned}
 T(Y \rightarrow X|Z) &= H(X, Z) - H(Z) + H(Y, Z) - H(X, Y, Z) \\
 &= - \sum_{i=1}^{d_X} \log \left(\frac{\alpha + i - 1}{\gamma_i^{(X)} \theta_i^{(X)}} \right) - \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + d_X + i - 1}{\gamma_i^{(Z)} \theta_i^{(Z)}} \right) \\
 &\quad + (\alpha + d_X + d_Z) [\Psi(\alpha + d_X + d_Z) - \Psi(\alpha)] \\
 &\quad - [\Psi(1) - \Psi(\alpha)] \left(d_X + d_Z - \sum_{i=1}^{d_X} \gamma_i^{(X)} - \sum_{i=1}^{d_Z} \gamma_i^{(Z)} \right) \\
 &\quad + \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + i - 1}{\gamma_i^{(Z)} \theta_i^{(Z)}} \right) - (\alpha + d_Z) [\Psi(\alpha + d_Z) - \Psi(\alpha)] \\
 &\quad + [\Psi(1) - \Psi(\alpha)] \left(d_Z - \sum_{i=1}^{d_Z} \gamma_i^{(Z)} \right) \\
 &\quad - \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + i - 1}{\gamma_i^{(Z)} \theta_i^{(Z)}} \right) - \sum_{i=1}^{d_Y} \log \left(\frac{\alpha + d_Y + i - 1}{\gamma_i^{(Y)} \theta_i^{(Y)}} \right)
 \end{aligned}$$

$$\begin{aligned}
 &+ (\alpha + d_Z + d_Y) [\Psi(\alpha + d_Z + d_Y) - \Psi(\alpha)] \\
 &- [\Psi(1) - \Psi(\alpha)] \left(d_Z + d_Y - \sum_{i=1}^{d_Z} \gamma_i^{(Z)} - \sum_{i=1}^{d_Y} \gamma_i^{(Y)} \right) \\
 &+ \sum_{i=1}^{d_X} \log \left(\frac{\alpha + i - 1}{\gamma_i^{(X)} \theta_i^{(X)}} \right) + \sum_{i=1}^{d_Z} \log \left(\frac{\alpha + d_X + i - 1}{\gamma_i^{(Z)} \theta_i^{(Z)}} \right) \\
 &+ \sum_{i=1}^{d_Y} \log \left(\frac{\alpha + d_X + d_Z + i - 1}{\gamma_i^{(Y)} \theta_i^{(Y)}} \right) \\
 &- (\alpha + d_X + d_Z + d_Y) [\Psi(\alpha + d_X + d_Z + d_Y) - \Psi(\alpha)] \\
 &+ [\Psi(1) - \Psi(\alpha)] \left(d_X + d_Z + d_Y - \sum_{i=1}^{d_X} \gamma_i^{(X)} - \sum_{i=1}^{d_Z} \gamma_i^{(Z)} - \sum_{i=1}^{d_Y} \gamma_i^{(Y)} \right) \tag{20}
 \end{aligned}$$

which, after simplifications, reduces to

$$\begin{aligned}
 T(Y \rightarrow X|Z) &= \sum_{i=1}^{d_Y} \log \left(1 + \frac{d_X}{\alpha + d_Z + i - 1} \right) \\
 &- d_Y \left[\sum_{i=1}^{d_X} \frac{1}{\alpha + i - 1} + \sum_{i=1}^{d_Z} \left(\frac{1}{\alpha + d_X + i - 1} - \frac{1}{\alpha + i - 1} \right) \right] \\
 &+ (\alpha + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_Z + i - 1} \\
 &- (\alpha + d_X + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_X + d_Z + i - 1} \tag{21}
 \end{aligned}$$

A.3. Multivariate Burr Distribution

The multivariate Burr density function for a d -dimensional random vector \mathbf{S} is:

$$f_{\mathbf{S}} = \prod_{i=1}^d (\alpha + i - 1) p_i c_i s_i^{c_i - 1} \left(1 + \sum_{j=1}^d p_j s_j^{c_j - 1} \right)^{-(\alpha + d)} \tag{22}$$

with $\mathbf{S} \in \mathbb{R}^d$, $s_i > 0$, $c_i > 0$, $d_i > 0$ for $i = 1, \dots, n$ and $\alpha > 0$. For the multivariate Burr distribution the differential entropy of \mathbf{S} is:

$$\begin{aligned}
 H(\mathbf{S}) &= - \sum_{i=1}^d \log(\alpha + i - 1) + (\alpha + d) [\Psi(\alpha + d) - \Psi(\alpha)] - \sum_{i=1}^d \log(c_i \sqrt[i]{p_i}) \\
 &+ [\Psi(\alpha) - \Psi(1)] \left(\sum_{i=1}^d \frac{c_i - 1}{c_i} \right) \tag{23}
 \end{aligned}$$

Thus, the transfer entropy for the multivariate Burr distribution can be formulated as:

$$\begin{aligned}
 T(Y \rightarrow X|Z) &= H(XZ) - H(Z) + H(YZ) - H(XYZ) \\
 &= - \sum_{i=1}^{d_X} \log(\alpha + i - 1) - \sum_{i=1}^{d_Z} \log(\alpha + d_X + i - 1)
 \end{aligned}$$

$$\begin{aligned}
 & + (\alpha + d_X + d_Z) [\Psi(\alpha + d_X + d_Z) - \Psi(\alpha)] \\
 & - \sum_{i=1}^{d_X} \log \left(c_i^{(X)} c_i^{(X)} \sqrt{p_i^{(X)}} \right) - \sum_{i=1}^{d_Z} \log \left(c_i^{(Z)} c_i^{(Z)} \sqrt{p_i^{(Z)}} \right) \\
 & + [\Psi(\alpha) - \Psi(1)] \left(\sum_{i=1}^{d_X} \frac{c_i^{(X)} - 1}{c_i^{(X)}} + \sum_{i=1}^{d_Z} \frac{c_i^{(Z)} - 1}{c_i^{(Z)}} \right) \\
 & + \sum_{i=1}^{d_Z} \log(\alpha + i - 1) - (\alpha + d_Z) [\Psi(\alpha + d_Z) - \Psi(\alpha)] \\
 & + \sum_{i=1}^{d_Z} \log \left(c_i^{(Z)} c_i^{(Z)} \sqrt{p_i^{(Z)}} \right) - [\Psi(\alpha) - \Psi(1)] \left(\sum_{i=1}^{d_Z} \frac{c_i^{(Z)} - 1}{c_i^{(Z)}} \right) \\
 & - \sum_{i=1}^{d_Z} \log(\alpha + i - 1) - \sum_{i=1}^{d_Y} \log(\alpha + d_Z + i - 1) \\
 & + (\alpha + d_Z + d_Y) [\Psi(\alpha + d_Z + d_Y) - \Psi(\alpha)] \\
 & - \sum_{i=1}^{d_Z} \log \left(c_i^{(Z)} c_i^{(Z)} \sqrt{p_i^{(Z)}} \right) - \sum_{i=1}^{d_Y} \log \left(c_i^{(Y)} c_i^{(Y)} \sqrt{p_i^{(Y)}} \right) \\
 & + [\Psi(\alpha) - \Psi(1)] \left(\sum_{i=1}^{d_Z} \frac{c_i^{(Z)} - 1}{c_i^{(Z)}} + \sum_{i=1}^{d_Y} \frac{c_i^{(Y)} - 1}{c_i^{(Y)}} \right) \\
 & + \sum_{i=1}^{d_X} \log(\alpha + i - 1) + \sum_{i=1}^{d_Z} \log(\alpha + d_X + i - 1) \\
 & + \sum_{i=1}^{d_Y} \log(\alpha + d_X + d_Y + i - 1) \\
 & - (\alpha + d_X + d_Z + d_Y) [\Psi(\alpha + d_X + d_Z + d_Y) - \Psi(\alpha)] \\
 & + \sum_{i=1}^{d_X} \log \left(c_i^{(X)} c_i^{(X)} \sqrt{p_i^{(X)}} \right) + \sum_{i=1}^{d_Z} \log \left(c_i^{(Z)} c_i^{(Z)} \sqrt{p_i^{(Z)}} \right) \\
 & + \sum_{i=1}^{d_Y} \log \left(c_i^{(Y)} c_i^{(Y)} \sqrt{p_i^{(Y)}} \right) \\
 & - [\Psi(\alpha) - \Psi(1)] \left(\sum_{i=1}^{d_X} \frac{c_i^{(X)} - 1}{c_i^{(X)}} + \sum_{i=1}^{d_Z} \frac{c_i^{(Z)} - 1}{c_i^{(Z)}} + \sum_{i=1}^{d_Y} \frac{c_i^{(Y)} - 1}{c_i^{(Y)}} \right) \tag{24}
 \end{aligned}$$

which, after simplifications, reduces to

$$\begin{aligned}
 T(Y \rightarrow X|Z) & = \sum_{i=1}^{d_Y} \log \left(1 + \frac{d_X}{\alpha + d_Z + i - 1} \right) \\
 & - d_Y \left[\sum_{i=1}^{d_X} \frac{1}{\alpha + i - 1} + \sum_{i=1}^{d_Z} \left(\frac{1}{\alpha + d_X + i - 1} - \frac{1}{\alpha + i - 1} \right) \right] \\
 & + (\alpha + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_Z + i - 1}
 \end{aligned}$$

$$-(\alpha + d_X + d_Z + d_Y) \sum_{i=1}^{d_Y} \frac{1}{\alpha + d_X + d_Z + i - 1} \tag{25}$$

B. Appendix

B.1. Multivariate Cauchy-Lorentz Distribution

The multivariate Cauchy-Lorentz density function for a d -dimensional random vector \mathbf{S} is:

$$f_{\mathbf{S}} = \frac{\Gamma(\frac{1+d}{2})}{\sqrt{\pi^{1+d}}} (1 + s_1^2 + s_2^2 + \dots + s_d^2)^{-\frac{1+d}{2}} \tag{26}$$

for $\mathbf{S} \in \mathbb{R}^d$. Interestingly, Equation (26) is equivalent to the multivariate t -distribution with one degree of freedom, zero expectation, and an identity covariance matrix [21]. For the case of $d = 1$, Equation (26) reduces to the univariate Cauchy-Lorentz density function [22]. The differential entropy of \mathbf{S} is:

$$H(\mathbf{S}) = -\log\left(\frac{\Gamma(\frac{1+d}{2})}{\sqrt{\pi^{1+d}}}\right) + \frac{1+d}{2} \left[\Psi\left(\frac{1+d}{2}\right) - \Psi\left(\frac{1}{2}\right) \right] \tag{27}$$

Thus, the transfer entropy $T(Y \rightarrow X|Z)$ for the multivariate Cauchy-Lorentz distribution can be formulated as:

$$\begin{aligned} T(Y \rightarrow X|Z) = & H(XZ) - H(Z) + H(YZ) - H(XYZ) \\ & - \log\left(\frac{\Gamma(\frac{1+d_X+d_Z}{2})}{\sqrt{\pi^{1+d_X+d_Z}}}\right) \\ & + \frac{1+d_X+d_Z}{2} \left[\Psi\left(\frac{1+d_X+d_Z}{2}\right) - \Psi\left(\frac{1}{2}\right) \right] \\ & + \log\left(\frac{\Gamma(\frac{1+d_Z}{2})}{\sqrt{\pi^{1+d_Z}}}\right) - \frac{1+d_Z}{2} \left[\Psi\left(\frac{1+d_Z}{2}\right) - \Psi\left(\frac{1}{2}\right) \right] \\ & - \log\left(\frac{\Gamma(\frac{1+d_Y+d_Z}{2})}{\sqrt{\pi^{1+d_Y+d_Z}}}\right) \\ & + \frac{1+d_Y+d_Z}{2} \left[\Psi\left(\frac{1+d_Y+d_Z}{2}\right) - \Psi\left(\frac{1}{2}\right) \right] \\ & + \log\left(\frac{\Gamma(\frac{1+d_X+d_Y+d_Z}{2})}{\sqrt{\pi^{1+d_X+d_Y+d_Z}}}\right) \\ & - \frac{1+d_X+d_Y+d_Z}{2} \left[\Psi\left(\frac{1+d_X+d_Y+d_Z}{2}\right) - \Psi\left(\frac{1}{2}\right) \right] \end{aligned} \tag{28}$$

which, after simplifications, using the identity in Equation (16), reduces to

$$\begin{aligned} T(Y \rightarrow X|Z) = & \log\left(\frac{\Gamma(\frac{1+d_Z}{2}) \Gamma(\frac{1+d_X+d_Y+d_Z}{2})}{\Gamma(\frac{1+d_X+d_Z}{2}) \Gamma(\frac{1+d_Y+d_Z}{2})}\right) + \frac{1+d_X+d_Z}{2} \xi\left(\frac{d_X+d_Z}{2}\right) \\ & - \frac{1+d_Z}{2} \xi\left(\frac{d_Z}{2}\right) + \frac{1+d_Y+d_Z}{2} \xi\left(\frac{d_Y+d_Z}{2}\right) \end{aligned}$$

$$- \frac{1 + d_X + d_Y + d_Z}{2} \xi \left(\frac{d_X + d_Y + d_Z}{2} \right) \quad (29)$$

where

$$\xi(a) = \sum_{i=1}^a \frac{1}{i - 0.5} \quad (30)$$

is obtained after a simplification of the digamma function.

References

1. Granger, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438.
2. Hlaváčková-Schindler, K.; Paluš, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **2007**, *441*, 1–46.
3. Guo, S.; Ladroue, C.; Feng, J. Granger Causality: Theory and Applications. In *Frontiers in Computational and Systems Biology*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 83–111.
4. Lock, J.G.; Jafari-Mamaghani, M.; Shafqat-Abbasi, H.; Gong, X.; Tyrcha, J.; Strömlad, S. Plasticity in the macromolecular-scale causal networks of cell migration. *PLoS One* **2014**, *9*, e90593.
5. Soofi, E.S. Principal information theoretic approaches. *J. Am. Stat. Assoc.* **2000**, *95*, 1349–1353.
6. Soofi, E.S.; Zhao, H.; Nazareth, D.L. Information measures. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 75–86.
7. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, doi:http://dx.doi.org/10.1103/PhysRevLett.85.461 .
8. Hlaváčková-Schindler, K. Equivalence of Granger causality and transfer entropy: A generalization. *Appl. Math. Sci.* **2011**, *5*, 3637–3648.
9. Seghouane, A.-K.; Amari, S. Identification of directed influence: Granger causality, Kullback-Leibler divergence, and complexity. *Neural Comput.* **2012**, *24*, 1722–1739.
10. Jafari-Mamaghani, M. Non-parametric analysis of Granger causality using local measures of divergence. *Appl. Math. Sci.* **2013**, *7*, 4107–4136.
11. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701.
12. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Multivariate Distributions, Models and Applications*; Volume 1; Wiley: New York, NY, USA, 2002.
13. Furman, E. On a multivariate gamma distribution. *Stat. Probab. Lett.* **2008**, *78*, 2353–2360.
14. Cover, T.M.; Thomas, J.A. *Elements of information theory*; Wiley: New York, NY, USA, 1991.
15. Florens, J.P.; Mouchart, M. A note on noncausality. *Econometrica* **1982**, *50*, 583–591.
16. Chow, G.C. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **1960**, *28*, 591–605.
17. Geweke, J. Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* **1982**, *77*, 304–313.
18. Ladroue, C.; Guo, S.; Kendrick, K.; Feng, J. Beyond element-wise interactions: Identifying complex interactions in biological processes. *PLoS One* **2009**, *4*, e6899.

19. Barrett, A.B.; Barnett, L.; Seth, A.K. Multivariate Granger causality and generalized variance. *Phys. Rev. E* **2010**, *81*, 041907.
20. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **2001**, *1*, 3–55.
21. Zografos, K.; Nadarajah, S. Expressions for Rényi and Shannon entropies for multivariate distributions. *Stat. Prob. Lett.* **2005**, *71*, 71–84.
22. Abe, S.; Rajagopal, A.K. Information theoretic approach to statistical properties of multivariate Cauchy-Lorentz distributions. *J. Phys. A* **2001**, *34*, doi:10.1088/0305-4470/34/42/301.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).