

Equivalence of Granger Causality and Transfer Entropy: A Generalization

Kateřina Hlaváčková-Schindler

Commission for Scientific Visualization
Austrian Academy of Sciences
Donau-City Str. 1, A-1220 Vienna, Austria
katerina.schindler@assoc.oeaw.ac.at

Abstract

Barnett et al. in 2009 proved that Granger causality and transfer entropy causality measure are equivalent for time series which have a Gaussian distribution. Granger causality test is linear, while transfer entropy a non-linear test. Many biological and physical mechanisms show to have non-Gaussian distributions. In this paper we investigate under which conditions on probability density distributions of the data can the equivalence of the two causality measures be extended. In the complexity sense "cheaper" linear Granger test can be applied for detection of causality in time series satisfying these conditions. These results have an impact on causality detection in common biological and physical time series.

Keywords: Granger causality, transfer entropy, Gaussian probability distribution, generalized Gaussian probability distribution

1 Introduction

Is Gaussian (normal) probability distribution sufficiently precise for real world data modeling? Dependence modeling by Gaussian probability distribution with copula functions has been widely used in applications of financial risk assessment and actuarial analysis, for example in the pricing of collateralized debt obligations. Some researches believe that the methodology of applying the Gaussian copula to credit derivatives to be one of the reasons behind the current global financial crisis. Gaussian distribution has been extensively applied also in modeling of neurophysiological time series, especially for EEG, see for instance [23], [9]. Normal children, in the early postnatal period, generate EEG's which have a non-Gaussian distribution of amplitude that becomes

increasingly Gaussian before one year of age and remains so throughout subsequent development. Conversely, the EEG's of children with Down's syndrome exhibit highly non-Gaussian properties at all ages studied [8].

On the other hand, results in the recent literature show that generalized Gaussian distributions often show better precision for modeling and classification of biomedical signal (EEG) than Gaussian distributions Generalized Gaussian distribution functions (GGD) for example [4].

Many biological mechanisms show log-normal distributions [17]. These processes in nature are for example log-normal: the length of latent periods of infectious diseases [21], distributions of mineral resources in the Earth's crust, the distribution of particles or chemicals, [7], reaction time, etc. However, some measurements can however on small data sample fit both normal and log-normal distributions (for example body height [19]). What is the difference between normal and log-normal distribution? Both forms of variability are based on a variety of forces (causes) acting independently of one another. A major difference is however that the effects can be additive or multiplicative, thus leading to normal or log-normal distributions, respectively [19]. Exponential distribution has been frequently uses in modeling in astrophysics, for example has been shown to be a good model for dusty galactic discs [20]. These probability distributions will be addressed in this paper.

By testing of causal relationships between time series, the appropriate model selection is of crucial importance. Barnett et al. in 2009 proved in [3] that the two measures for testing the causal relationships between multivariate series, namely G-causality and transfer entropy, are equivalent for time series which have a Gaussian distribution. Taking in consideration that measurements observable in nature are often non-Gaussian, it opens a question, for which probability distributions, could be this result extended. In this paper we will investigate under which conditions on probability density distributions of the data can be the equivalence of the two causality measures extended.

1.1 Mixtures of Gaussian distributions and one generalized Gaussian distribution for time series modeling

Hidden Markov Models (HMM) with Gaussian mixtures have been frequently used in literature for sequential data classification, including the EEG and other biomedical signals. The number of Gaussian mixtures is usually selected ad hoc, which influences the quality of modeling. Instead of using Gaussian mixtures, Bicego et al. in [4] suggested to apply a Hidden Markov Model where each state dependent emission probability function is represented by one Generalized Gaussian (GG). Although modeling each emission function as a mixture of GGs is also possible (as is typically done for standard Gaussians),

they restricted the formulation to HMMs with just one GG per state. It has been shown in [5] that given one HMM using a mixture of Gaussians in each state, there exists an (in a likelihood sense) equivalent HMM with more states but just one Gaussian per state, and this proof could be easily extended to any kind of mixture. Using one GG per state eliminates the problem of choosing the number of components in each mixture. In the EEG signal classification, a slightly better performance of HMM with one generalized Gaussian has been experimentally shown over HMM with one Gaussian distribution function in [4]. For multivariate Generalized Gaussian Hidden Markov Models (GG-HMM) a remarkable improvement in the classification over multivariate Gaussian Hidden Markov Models (G-HMM) has been achieved, similarly also in [12]. These experimental conclusions motivate the application of multivariate generalized Gaussian distributions for biomedical data.

2 Causality testing in time series

The introduction of the concept of causality into the experimental practice, namely into analyses of data observed in consecutive time instants, time series, is due to Clive W. J. Granger in 1969 in [10], the 2003 Nobel prize winner in economy. In his Nobel lecture [11] he recalled the inspiration by the Wiener's work and identified two components of the statement about causality: 1. The cause occurs before the effect; and 2. The cause contains information about the effect that is unique, and is in no other variable. As Granger put it, a consequence of these statements is that the causal variable can help to forecast the effect variable after other data has been first used [11]. This restricted sense of causality, referred to as *Granger causality*, GC thereafter, characterizes the extent to which a process X_t is leading another process Y_t , and builds upon the notion of incremental predictability. It is said that the *process X_t Granger causes another process Y_t* if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t . The standard test of GC developed by Granger [10] is based on a linear regression model.

2.1 Granger causality

We will adopt the notation of the paper from Barnett et al [3]. Let \oplus denotes concatenation of vectors, so that for $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{y} = (y_1, \dots, y_m)$ $\mathbf{x} \oplus \mathbf{y}$ is the $1 \times (d+m)$ vector $(x_1, \dots, x_d, y_1, \dots, y_m)$. Given jointly distributed multivariate random variables \mathbf{X} and \mathbf{Y} i.e. random vectors in R^d , we denote by $\Sigma(\mathbf{X})$ the $d \times d$ matrix of covariances $cov(X_i, Y_j)$ and by $\Sigma(\mathbf{X}, \mathbf{Y})$ the $d \times m$

matrix of cross-covariances $\text{cov}(X_i, Y_\alpha)$. Let $\Sigma(\mathbf{X}|\mathbf{Y})$ denotes the $d \times d$ matrix

$$\Sigma(\mathbf{X}|\mathbf{Y}) = \Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{Y})\Sigma(\mathbf{Y})^{-1}\Sigma(\mathbf{X}, \mathbf{Y})^T \quad (1)$$

define when $\Sigma(\mathbf{Y})$ is invertible.

Suppose we have a stationary multivariate stochastic process \mathbf{X}_t in discrete time (i.e. marginal distributions are jointly distributed). Denote $\mathbf{X}_t^{(p)} = \mathbf{X}_t \oplus \mathbf{X}_{t-1} \oplus \dots \oplus \mathbf{X}_{t-p+1}$ for \mathbf{X} along with $p-1$ lags so that $\mathbf{X}_t^{(p)}$ is a $1 \times pd$ random vector for each t . Given the lag p , we use the shorthand notation $\mathbf{X}_t^- = \mathbf{X}_{t-1}^{(p)}$ for the lagged variable.

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

$$\mathbf{X}_t = \alpha_t + (\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Z}_{t-1}^r).A + \epsilon_t \quad (2)$$

$$\mathbf{X}_t = \alpha'_t + (\mathbf{X}_{t-1}^{(p)} \oplus \mathbf{Y}_{t-1}^q \oplus \mathbf{Z}_{t-1}^r).A' + \epsilon'_t \quad (3)$$

where A and A' are the matrices of regression coefficients, α_t and α'_t are the constant terms and the random vectors ϵ and ϵ' comprise the residuals, so that so that the predictee variable \mathbf{X} is regressed firstly on the previous p lags of itself plus r lags of the conditioning variable \mathbf{Z} and secondly, in addition, on q lags of the predictor variable \mathbf{Y} . The G-causality of \mathbf{Y} to \mathbf{X} given \mathbf{Z} is a measure of the extent to which inclusion of \mathbf{Y} in the second model (3) reduces the prediction error of the first model (2). The standard measure of G-causality in the literature is defined for univariate predictor and predictee variables \mathbf{Y} and \mathbf{X} , and is given by the natural logarithm of the ratio of the residual variance in the restricted regression (2) to that of the unrestricted regression (3). It was shown in [3] that for the G-causality holds

$$\mathcal{F}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}} = \ln\left(\frac{\Sigma(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Z}^-)}{\Sigma(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-)}\right). \quad (4)$$

By stationarity this expression does not depend on time t , so we omit t from the notation.

2.2 Transfer entropy

Transfer entropy as a non-linear causality measure was introduced by Schreiber in 2000 [22]. Let us first remind some basic definitions. The differential entropy of a (continuous) random vector \mathbf{X} taking its values in R^d with the probability density function $p(\mathbf{x})$ is defined by

$$h(\mathbf{X}) = - \int_{R^d} p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}.$$

If \mathbf{X} is a discrete (multivariate) random variable given by a set of possible values $\{x_1, \dots, x_n\}$ then the entropy can explicitly be written as

$$H(\mathbf{X}) = - \sum_{i=1}^n p(\mathbf{x}_i) \ln p(\mathbf{x}_i)$$

where p denotes the probability mass function of X . With $\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t$ defined as before, the transfer entropy of \mathbf{Y} to \mathbf{X} given \mathbf{Z} is defined as the difference between the entropy of X conditioned on its own past and the past of \mathbf{Z} , and its entropy conditioned, in addition, on the past of \mathbf{Y} :

$$\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X} | \mathbf{Z}} = H(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Z}^-) - H(\mathbf{X} | \mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-) \quad (5)$$

where $H(\cdot | \cdot)$ is the conditional entropy. For stationary variables the transfer entropy does not depend on t , so we omitted it from labeling. More on the information-theoretic methods for causality detection can be found in our review paper [13]. As it was already mentioned in the Introduction, Barnett et al. proved in [3] that if all processes are jointly Gaussian then Granger causality and transfer entropy are equivalent up to a factor of 2. This result provides for the first time a unified framework for data-driven causal inference that bridges information-theoretic and autoregressive methods. This statement brings a consequence for practice, a reduction of the computational complexity. In the complexity sense "cheaper" linear test can be applied for detection of causality, when one knows the time series in Gaussian.

In our paper we investigate, to which other multivariate probability distributions can be the equivalence of the two causality measures extended. To express the transfer entropy, we need the knowledge of the analytical value of the particular entropy in multivariate case. In the next session we discuss common multivariate joint distributions, for which are differential entropies in their analytical form known.

3 Jointly multivariate probability distributions and their differential entropies

Recall that a jointly multivariate probability distribution f of dimension d can be defined as

$$f_{X_1, \dots, X_d}(x_1, \dots, x_d) = f_{X_1}(x_1) f_{X_2 | X_1}(x_2 | x_1) \dots f_{X_d | X_1 \dots X_{d-1}}(x_d | x_1 \dots x_{d-1})$$

where

$$f_{X_i | X_1 \dots X_{i-1}}(x_i | x_1 \dots x_{i-1}) = \frac{\int \dots \int f_{X_1, \dots, X_d}(x_1, \dots, x_i, u_{i+1}, \dots, u_d) du_{i+1} \dots du_d}{\int \dots \int f_{X_1, \dots, X_d}(x_1, \dots, x_{i-1}, u_i, \dots, u_d) du_i du_{i+1} \dots du_d}$$

$$f_{X_1, \dots, X_i}(x_1, \dots, x_i) = \int \dots \int f_{X_1, \dots, X_d}(x_1, \dots, x_i, x_{i+1}, \dots, x_d) dx_{i+1} \dots dx_d.$$

The definition of the multivariate probability distribution by copula theory is not considered here. Analytical expressions for the entropy of various univariate continuous distributions are known (i.e. Lazo and Rathie [18]) but their extension to multivariate distributions is not always trivial. The formulas for various common multivariate distributions were computed by Darbellay and Vajda in [6], namely for d -dimensional Pareto, logistic Burr, exponential, Weibull, Weimann exponential, Ordered Weimann exponential and Gamma-exponential distributions.

3.1 Normal distribution and its entropy

Recall that a multivariate normal distribution in R^d is defined as

$$N(\mathbf{x}, \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)$$

where $|\Sigma|$ is the determinant of Σ , a $d \times d$ symmetric positive definite matrix. The entropy of this distribution is $\ln \sqrt{(2\pi e)^d |\Sigma|}$.

3.2 Generalized-normal distribution and symmetric Kotz type distribution

The definitions of this subsection are adopted from Kitsos and Toulas in [16]. Both presented distributions are generalizations of a normal distribution. The d -dimensional random variable X has the γ -order **generalized normal distribution** with mean μ and covariance matrix Σ when the density function is of the form

$$KT_\gamma^d(\mu, \Sigma) = C_{d,\gamma} |\Sigma|^{-1/2} \exp\left\{-\frac{\gamma-1}{\gamma} Q(\mathbf{X})^{\frac{\gamma}{2(\gamma-1)}}\right\} \quad (6)$$

where $|\Sigma|$ means determinant of Σ , $Q(\mathbf{X}) = \langle (\mathbf{X} - \mu)^T, \Sigma^{-1}(\mathbf{X} - \mu) \rangle$, where $\langle u^T, v \rangle$ is the inner product of $u, v \in R^d$ and T denotes transpose. The normalizing factor is defined as

$$C_{d,\gamma} = \pi^{-\frac{d}{2}} \frac{\Gamma(\frac{d}{2} + 1)}{\Gamma(d\frac{\gamma-1}{\gamma} + 1)} \left(\frac{\gamma-1}{\gamma}\right)^{d\frac{\gamma-1}{\gamma}} \quad (7)$$

where $\Gamma(\cdot)$ is the Gamma function. For $\gamma = 2$ is $KT_2(\mathbf{x}, \mu, \Sigma)$ is the (multivariate) normal distribution.

One of the merits of the generalized normal distributions is that they belong to the Kotz-type distribution family (see below), i.e., they are elliptically

contoured distributions. Recall that the symmetric Kotz type distribution has density

$$Kotz_{m,r,s}(\mu, \Sigma) = K(m, r, s) |\Sigma|^{-1/2} Q^{m-1} \exp\{-rQ^s\}$$

where $r > 0, s > 0, 2m + d > 2$ and the normalizing constant $K(m, r, s)$ is given by:

$$K(m, r, s) = \frac{s\Gamma(d/2)r^{(2m+d-2)/2s}}{\pi^{d/2}\Gamma(\frac{2m+d-2}{2s})}$$

It was shown in [14] that the distribution $KT_\gamma(\mu, \Sigma)$ is the symmetric Kotz type distribution with parameters $m = 1, s = \frac{\gamma}{2(\gamma-1)}, r = \frac{\gamma-1}{\gamma}$, i.e. $KT_\gamma(\mu, \Sigma) = K_{1, \frac{\gamma-1}{\gamma}, \frac{\gamma}{2(\gamma-1)}}(\mu, \Sigma)$.

Note also that for the normal distribution holds $N(\mu, \Sigma) = KT_2(\mu, \Sigma) = Kotz_{1, \frac{1}{2}, 2}(\mu, \Sigma)$, while the normalizing factor is $C_{d,\gamma} = K(1, \frac{\gamma}{2(\gamma-1)}, \frac{\gamma-1}{\gamma})$. It has been proven in [2] that the entropy of the symmetric Kotz type distribution $K_{m,r,s}(\mu, \Sigma)$ is $H(Kotz_{m,r,s}(\mu, \Sigma)) = -\ln C(d, s, m) + \frac{1}{2} \ln |\Sigma| + \frac{2m-d+2}{2s} - \frac{m-1}{s} [\psi(\frac{2m+d-2}{2s}) - \ln r]$, where $\psi(\mathbf{x}) = \frac{d}{d\mathbf{x}}[\ln \Gamma(\mathbf{x})]$.

By asserting $m = 1, s = \frac{\gamma}{2(\gamma-1)}, r = \frac{\gamma-1}{\gamma}$, the entropy of generalized normal multivariate distribution was computed in [15] as

$$H(K_{1, \frac{1}{2}, 2}(\mu, \Sigma)) = -\ln \frac{|\Sigma|^{1/2}}{C_{d,\gamma}} + d \frac{\gamma-1}{\gamma} \tag{8}$$

4 Results

Lemma 4.1. *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be jointly multivariate random discrete variables with values in R^d . Assume that the entropy of their distributions can be expressed as $H(\mathbf{X}) = C \ln(|\Sigma(\mathbf{X})|) + S(d)$, where C is independent of \mathbf{X} and d and for $S : Z^+ \rightarrow R$ holds $S(d) + S(p) = S(d + p)$. Then the Granger causality and transfer entropy for the discrete variables are equivalent up to the multiplication constant $\frac{1}{C}$.*

Proof:

The proof uses some ideas of the proof from [3] for Gaussian distributions. If a (jointly multivariate) continuous probability distribution \mathbf{X} has entropy in the form $h(\mathbf{x}) = \mathbf{C} \ln(\mathbf{x}) + \mathbf{S}$ then its discrete entropy can be written as $H(\mathbf{X}) = C \ln(|\Sigma(\mathbf{X})|) + S$ where Σ is a covariance matrix. It follows from the block determinant identity $\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D||A - BD^{-1}C|$ and the definition of Σ that $\Sigma(\mathbf{X} \oplus \mathbf{Y}) = \begin{pmatrix} \Sigma(\mathbf{X}) & \Sigma(\mathbf{X}, \mathbf{Y}) \\ \Sigma(\mathbf{X}, \mathbf{Y})^T & \Sigma(\mathbf{Y}) \end{pmatrix} = |\Sigma(\mathbf{Y})| |\Sigma(\mathbf{X}) \Sigma(\mathbf{X}, \mathbf{Y}) \Sigma(\mathbf{Y})^{-1} \Sigma(\mathbf{X}, \mathbf{Y})^T|$ which from formula (1) equals to $|\Sigma(\mathbf{Y})| \cdot |\Sigma(\mathbf{X}|\mathbf{Y})|$. Then $H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X} \oplus$

$\mathbf{Y}) - \mathbf{H}(\mathbf{Y}) = C \ln(|\Sigma(\mathbf{X} \oplus \mathbf{Y})|) - \mathbf{H}(\mathbf{Y}) = C \ln(|\Sigma(\mathbf{Y})||\Sigma(\mathbf{X}|\mathbf{Y})|) + S(2d) - C \ln |\Sigma(\mathbf{Y})| - S(d) = C \ln(\frac{|\Sigma(\mathbf{Y})||\Sigma(\mathbf{X}|\mathbf{Y})|}{|\Sigma(\mathbf{Y})|}) + S(d) = C \ln(|\Sigma(\mathbf{X}|\mathbf{Y})|) + S(d).$

Analogously we get $H(\Sigma(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Z}^-)) = C \ln |\Sigma(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Z}^-)| + S(d)$ and $H(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-) = C \ln |\Sigma(\mathbf{X}|\mathbf{X}^- \oplus \mathbf{Y}^- \oplus \mathbf{Z}^-)| + S(d)$. The difference of the two last formulas equals to transfer entropy $\mathcal{T}_{\mathbf{Y} \rightarrow \mathbf{X}|\mathbf{Z}}$ and comparing it to the G-causality defined by (4), we get the statement of the lemma.

4.1 Generalized normal distribution

In the following theorem we set conditions under which are G-causality and transfer entropy equivalent.

Theorem 4.2. *Let random discrete variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ have jointly multivariate generalized normal distributions in R^d whose continuous version is given by formula (6). If the parameter $\omega = \frac{\gamma-1}{\gamma}$ of the distribution satisfies the condition*

$$\omega d = \ln\left[\frac{\Gamma(\frac{3d}{2})\Gamma(2d\omega)\omega^{\omega d}}{\pi^{\frac{d}{2}}\Gamma(3d\omega)\Gamma(d)}\right], \tag{9}$$

then the Granger causality and transfer entropy of these variables are equivalent up to a factor of 2.

Proof:

Denote by $H_d = H(x_1, \dots, x_d)$ the entropy of a d -dimensional random variable. From the definition of entropy (8) we get $H_{2d} - H_{3d} = \ln |\Sigma_{2d}|^{1/2} - \ln |\Sigma_{3d}|^{1/2} - \ln C_{2d,\gamma} + 2d\frac{\gamma-1}{\gamma} + \ln C_{3d,\gamma} - 3d\frac{\gamma-1}{\gamma}$. To apply Lemma 4.1, one requires that $-\ln C_{2d,\gamma} + 2d\frac{\gamma-1}{\gamma} + \ln C_{3d,\gamma} - 3d\frac{\gamma-1}{\gamma} = 0$. This corresponds to $\frac{C_{3d,\gamma}}{C_{2d,\gamma}} = \exp^{d\frac{\gamma-1}{\gamma}}$. This rewritten equals to

$$\frac{\pi^{-\frac{3d}{2}}\Gamma(\frac{3d}{2} + 1)\omega^{3d\omega}\Gamma(2d\omega)}{\Gamma(3d\omega + 1)\pi^{-d}\Gamma(d + 1)(\omega)^{2d\omega}} = \exp^{d\omega} .$$

Applying the fact that $\Gamma(z + 1) = z\Gamma(z)$ for every $z \in R$, we get

$$\frac{\pi^{-\frac{d}{2}}(\omega)^{d\omega}\frac{3d}{2}\Gamma(\frac{3d}{2})2d(\omega)\Gamma(2d\omega)}{(3d\omega)\Gamma(3d\omega).d.\Gamma(d)} = \exp^{d\omega}$$

$$\frac{\Gamma(\frac{3d}{2})\Gamma(2d\omega)(\omega)^{d\omega}\frac{3d}{2}2d\omega}{\pi^{\frac{d}{2}}\Gamma(3d\omega)\Gamma(d)(3d\omega)d} = \exp^{d\omega}$$

$$\frac{\Gamma(\frac{3d}{2})\Gamma(2d\omega)(\omega)^{d\omega}}{\pi^{\frac{d}{2}}\Gamma(3d\omega)\Gamma(d)} = \exp^{d\omega}$$

and the statement of the theorem follows.

From the graphical solution of equation (9) we conclude that it is fulfilled for every $1 < d \leq 77$. For higher dimensions is the argument of logarithm not positive. It is of a principal interest, for which $C_{d,\gamma}$ would be the corresponding generalized Gaussian distribution suitable for modeling the common biological time series, namely EEG time series. This question can be answered by direct application of these distributions to the EEG data.

4.2 Other probability distributions

The equivalence of Granger causality and transfer entropy holds also for the following probability distributions.

Theorem 4.3. 1. Let random discrete variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ have jointly multivariate Weibull exponential distributions in R^d whose continuous density has the form

$$p(\mathbf{x}) = \prod_{i=0}^{d-1} \frac{1}{\theta_i} e^{-\frac{x_{i+1}-x_i}{\theta_i}}$$

where $x_i > 0$ are arranged in increasing order of magnitude, with $\theta_i > 0$ for $i = 1, \dots, d-1$. Then the Granger causality and transfer entropy of these variables are equivalent up to the factor of 1.

2. Let random discrete variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ have jointly multivariate log-normal distributions in R^d with the continuous density in the form

$$p(\mathbf{x}) = (2\pi)^{-d/2} |\Sigma|^{-1/2} (\prod_{i=1}^d x_i)^{-1} \exp -\frac{1}{2} (\log \mathbf{x} - \mu)^T \Sigma^{-1} \log(\mathbf{x} - \mu)$$

with $\log \mathbf{x} = (\log x_1, \dots, \log x_d)^T$, $x_i > 0, i = 1, \dots, d$, $\mu \in R^d$ and Σ a positive definite matrix of order d . Then their Granger causality and transfer entropy are equivalent up to the factor of $\frac{1}{2}$.

Proof: (i): It has been shown in [6] and [24] that the entropy of multivariate Weibull exponential distribution is $H(p) = \sum_{i=0}^{d-1} \log \theta_i + d$. This entropy for discrete distributions expressed by means of $|\Sigma(X)|$ fulfills our lemma and the statement follows.

(ii): It has been shown in [1] and [24] that the entropy of multivariate log normal distribution is $H(p) = \frac{1}{2} \log |\Sigma| + \frac{d}{2} + \frac{d}{2} \log(2\pi) + \sum_{i=1}^d \mu_i$. This entropy for discrete distributions rewritten by means of $|\Sigma(X)|$ fulfills our lemma and the statement follows.

5 Conclusion

Motivated by the experimental evidence, that the generalized normal distribution is more suitable for EEG time series modeling than the normal distribution, we have investigated what influence on the causality detection these models have. We have got a condition for the parameters of a generalized normal distribution, for which the equivalence of the Granger causality and transfer entropy holds up to a factor. For exponential Weibull and log-normal data distribution we have also proven the equivalence of the two causality measures. These probability distributions also occur in natural processes. In the complexity sense "cheaper" linear test can be applied for detection of causality in time series, when the probability distributions are one of the mentioned ones. The next step of our research is to investigate experimentally, how precise do the log-normal, Weibull distributions and the generalized normal distributions satisfying the condition (9), respectively model the EEG time series or other real world time series.

References

- [1] N.A. Ahmed and D.V. Gokhale, Entropy expressions and their estimators for multivariate distributions, *IEEE Trans. Inform. Theory*, 35 (1989), 688-692.
- [2] G. Aulogiaris and K. Zografos, A maximum entropy characterization of symmetric Kotz type and Burr multivariate distributions, *Sociedad de Estadística e Investigación Operativa, Test*, 13, 1 (2004), 65-83.
- [3] L. Barnett, A.B. Barrett and A.K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables, *Phys. Rev. Lett.* 103, 23 (2009).
- [4] M. Bicego, D. Gonzalez-Jimenez, E. Grosso and J.L. Alba Castro, Generalized Gaussian distributions for sequential data classification, *International Conference on Pattern Recognition* (2008), 1-4.
- [5] M. Bicego, V. Murino and M. Figueiredo, A sequential pruning strategy for the selection of the number of states in Hidden Markov Models, *Pattern Recognition Letters*, 24, 9-10 (2003) 1395-1407.
- [6] G.A. Darbellay and I. Vajda, Entropy expressions for multivariate continuous distributions, *IEEE Trans. Inform. Theory*, 46, 2 (2000), 709-712.

- [7] C. Di Gorgio, A. Krempff, H. Guiraud, P. Binder, C. Tiret and G. Dumenil, Atmospheric pollution by airborne microorganisms in the City of Marseilles, *Atmospheric Environment* 30 (1996), 155-160.
- [8] R. Elul, J. Hanley and J.Q. Simmons, Non-Gaussian behavior of the EEG in Down's syndrome suggests decrease neural connections, *Acta Neurologica Scandinavica*, Vol. 51, Issue 1 (1975), 21-28.
- [9] S. Faul, G. Gregorcic, G. Bolyan, W. Marnane, G. Lightbody and S. Conolly, Gaussian process modeling of EEG for the detection of neonatal seizures, *IEEE Trans. Biomed Eng.* 54, 12 (2007), 2151-62.
- [10] C.W.J. Granger, Investigating causal relations by econometric and cross-spectral methods, *Econometrica*, 37 (1969), 424-438.
- [11] C.W.J. Granger, Time series analysis, cointegration, and applications. Nobel Lecture, December 8, 2003.
- [12] C.W. Hesse, D. Holtackers and T. Heskes, On the use of mixtures of Gaussian and generalized exponentials for modelling and classification of biomedical signals, Belgian Day on Biomedical Engineering December 7-8, 2006, IEEE Benelux EMBS Symposium.
- [13] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka and J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, *Physics Reports* 441 (2007), 1 - 46.
- [14] C.P. Kitsos and N.K. Tavoularis, On the hyper-multivariate normal distribution, Interantional Workshop on Applied Probabilities, Paris, July 7-10 (2008).
- [15] C.P. Kitsos and N.K. Tavoularis, New Entropy Type Information Measures, Proceedings of the ITI 2009 , 31st Int. Conf. on Information Technology Interfaces, Vavtat, Croatia, June 22-25 (2009).
- [16] C.P. Kitsos and T.L. Toulas, New Information Measures for the Generalized Normal Distribution, *Information*, 1 (2010), 13-27.
- [17] A.L. Koch, The logarithm in biology. I. Mechanisms generating the log-normal distribution exactly. *Journal of Theoretical Biology*, 23 (1969), 251-268.
- [18] A.C.G.V. Lazo and P.N. Rathie, On the entropy of continuous probability distributions, *IEEE Trans. Inform. Theory*, 24 (1978), 120-122.

- [19] E. Limpert, W.A. Stahel and M. Abbt, Log-normal Distributions across the Sciences: Keys and Clues, *BioScience*, 51, 5 (2001), 341-352.
- [20] A. Misiriotis, N.D. Kylafis, J. Papamastorakis and E.M. Xilouris, Is the exponential distribution a good approximation of dusty galactic disks? *Astron. Astrophys.* 353 (2000), 117123.
- [21] P.E. Sartwell, The distribution of incubation periods of infectious disease. *America Journal of Hygiene* 51 (1950), 310-318.
- [22] T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.* 85, 2000, 461-464.
- [23] M. Zhong, F. Lotte, M. Girolami and A. Lécuyer, Classifying EEG for brain computer interfaces using Gaussian processes, *Pattern Recognition Letters*, 29, 3 (2008), 354-359.
- [24] K. Zografos, S. Nadarajah, Expressions for Renyi and Shannon entropies for multivariate distributions. *Statistics and Probability Letters*, 71 (2005), 71-84.

Received: May, 2011