Comparing Quantization Methods for Testing Causality between Continuous-state and Discrete-time Processes

Juliana M. de Assis, Saulo O. D. Luiz, Francisco M. de Assis

***Abstract*— This paper investigates the use of the plug-in directed information rate estimator and hypothesis testing to detect causality between discrete-time continuous-state processes. Following our previous work, we have first quantized the continuous-state processes by means of three quantization methods: equidistant, equipopulated and symbolic. Then, we have applied directed information rate estimation and hypothesis testing. Simulations indicate that the symbolic method is the most effective in discriminating cases with an underlying causality from those with abscent causality.**

***Keywords*— Directed information, Causality, Estimation, Quantization, Hypothesis testing.**

I. Introduction

Causality is one fundamental concept in science. According to Wiener [1], a process $\{X_n\}$ causes a process $\{Y_n\}$ if knowledge about the past of $\{X_n\}$ improves prediction of $\{Y_n\}$, instead of using prediction from the past of $\{Y_n\}$ alone [2]. Information theory tackles the issue of identifying causal relations among processes by introducing measures such as directed information (DI), transfer entropy, conditional mutual information and momentary information transfer [3]. Despite some critiques on the use of these measures [4], DI in particular has been widely used in diverse areas, such as economy [5], neuroscience [6], [7], [8], hydrology [3] and control [9].

When detecting causal influences in real data by means of DI, usually one needs to estimate DI, because the underlying probability measures are unknown. Thus, considerable effort has been made to develop DI estimators. There are DI parametric estimators for discrete-state processes, as proposed by Quinn *et al*. [6], and non-parametric estimators, as proposed by Jiao *et al*. [10]. There are also DI estimators for continuous-state processes [11], as proposed by Murin *et al*., which extend mutual information estimators [12]. There is also a DI estimator for the case when one process is a discrete-state process and the other is a continuous-state process [13]. In our previous work, we evaluated DI estimates obtained from quantization of continuous-state processes by means of three different methods followed by the use of Jiao estimators [14].

Despite accurate estimates obtained from many of these estimators, some of them are rather complex and computationally demanding. Moreover, it may be difficult to interpret the meaning of a DI estimate and whether it means a causal relation indeed. The work of Kontoyiannis and Skoularidou [15], however, provides evidence that the plug-in estimator (maximum likelihood estimator) allows the use of statistical tests for detecting causality among Markov chains (under broad assumptions). Motivated by this achievement, in this paper we investigate whether quantization of continuous-state processes followed by plug-in estimation and hypothesis testing allows detection of causality. In order to do so, we use examples as a ground truth and use Monte Carlo simulation to assess this approach.

This paper is organized as follows: Section II introduces notation and terminology used in this paper and Section III introduces the measures of DI and DI rate. Section IV presents the methods used in this paper, and Section V presents the obtained results. Finally, Section VI concludes the paper.

Juliana M. de Assis, Saulo O. D. Luiz and Francisco M. de Assis, Department of Electrical Engineering, Federal University of Campina Grande, Campina Grande-PB, Brazil, e-mail: juliana.assis@ee.ufcg.edu.br . This work was partially supported by CAPES.

II. Notation and Terminology

In this paper, we denote random variables by uppercase letters, continuous-state random processes by uppercase letters in braces and with subscript n (for example, $\{X_n\}$). We denote discrete-state random processes similarly, but using a tilde (e.g. $\{\tilde{X}_n\}$). Subscripts usually denote the variable index in a sequence, for example, X_{11} generally indicates the 11th component of the process $\{X_n\}$. Subscripts also denote the first index considered in a sequence. Superscripts on a random variable denote finite length sequences of this random variable, for example, $X_2^5 = (X_2, \ldots, X_5)$. Subscript omission in a sequence means that it begins in index 1 ($X^5 = (X_1, X_2, \ldots, X_5)$). Throughout this paper, $\log$ is base 2 and $\mathbf{I}_{\{A=a\}}$ is the indicator function of the expression in subscript ($\{A = a\}$ in this example). The alphabet of a discrete-state random process $\{\tilde{X}_n\}$ is given in caligraphic letters (as $\mathcal{X}$) and its cardinality is given as $|\mathcal{X}|$. Moreover, $P_{\tilde{X}}(x)$ is the probability of the event $\{\tilde{X} = x\}$ and $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2.

III. Directed Information

As mentioned in our previous work [14], DI between the finite duration sequences assuming continuous values X^n and Y^n is defined as [16]:

$$I(X^n \to Y^n) \quad = \quad h(Y^n) - h(Y^n||X^n) \tag{1}$$

where

$$h(Y^n) = \sum_{i=1}^{n} h(Y_i|Y^{i-1}) \tag{2}$$

is the differential entropy of the sequence Y^n (and in the right side of (2) we have applied the chain rule for differential entropy) [17] and

$$h(Y^n||X^n) = \sum_{i=1}^{n} h(Y_i|Y^{i-1}X^i) \tag{3}$$

is the causally conditioned differential entropy of sequence Y^n causally conditioned on X^n.

Usually there is interest in how DI increases as the sequence grows, or in its rate - DI per letter, which is defined as [18]:

$$I_n(X \to Y) = \frac{1}{n} I(X^n \to Y^n). \tag{4}$$

DI rate is also defined in the limit as $n \to \infty$, when the limit exists:

$$I_\infty(X \to Y) = \lim_{n\to\infty} I_n(X \to Y). \tag{5}$$

There are analogous definitions when dealing with discrete-state random processes.

IV. Methods

In this Section we explain the methods used in this paper. Since the plug-in DI estimator is suitable to discrete-state processes, quantization is necessary before using it with continuous-state processes.

A. Quantization Methods

Previously to plug-in estimation of DI and hypothesis testing, it is necessary to quantize the amplitudes of the continuous values. The quantization methods adopted here are equidistant, equipopulated and symbolic [19],[20].

Both equidistant and equipopulated methods segment the range of each sample function, and the quantized process $\{\tilde{X}_n\}$ assumes values $\tilde{X}_i = j$ if X_i is in the j-th segment. The equidistant method partitions the range in equidistant segments, according to Euclidean distance, while the equipopulated method partitions the range in equipopulated segments (it is a classical histogram method [19]).

The symbolic method follows the work of Bandt and Pompe [20]. It consists in ordering k consecutive values of the time series, representing this transition order by one number that represents one of $k!$ possible permutations. For example, consider the time series:

$$X^7 = [0.5\ 0.75\ -0.1\ -0.23\ 0.05\ 0.52\ 0.49]$$

If we choose $k = 2$, we have the corresponding transitions:

$$[01\ 10\ 10\ 01\ 01\ 10].$$

Labelling the symbol 01 as 1 and the symbol 10 as 2, we obtain the following quantized sequence (beginning in index $n = 2$):

$$\tilde{X}_2^7 = [1\ 2\ 2\ 1\ 1\ 2].$$

On the other hand, if we choose $k = 3$ we have the corresponding transitions:

$$[120\ 210\ 102\ 012\ 021].$$

TABLE I
PERMUTATION VALUES

Transition	012	021	102	120	201	210
Discretized value	1	2	3	4	5	6

Labelling each transition as in Table I, we obtain the following quantized sequence (beginning in index $n = 3$):

$$\tilde{X}_3^7 = [4\ 6\ 3\ 1\ 2].$$

B. DI plug-in estimation and hypothesis testing

Plug-in estimation basically counts the occurrence of each value of the alphabet and assigns the relative frequency to the probability of each value. Also, these relative frequencies are used in the functional of DI. Relative frequencies are the maximum likelihood estimator of a probability vector of a variable with a discrete alphabet [21]. Moreover, if the joint process $\{(\tilde{X}_n, \tilde{Y}_n)\}$ is an irreducible and aperiodic Markov chain of finite order, then the plug-in DI rate estimator is consistent [15].

Formally, the plug-in DI rate estimate is given as follows. Given a sample $(\tilde{X}_{-k+1}^n, \tilde{Y}_{-k+1}^n)$ of the joint process $\{(\tilde{X}_n, \tilde{Y}_n)\}$, the relative frequency of the occuring sequence (a_0^k, b_0^k) in the sample is [15]:

$$\hat{P}_{\tilde{X}_{-k}^0 \tilde{Y}_{-k}^0, n}(a_0^k, b_0^k) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{\{\tilde{X}_{i-k}^i = a_0^k, \tilde{Y}_{i-k}^i = b_0^k\}}. \tag{6}$$

Thus, the DI rate estimate is:

$$\hat{I}_n^{(k)}(\tilde{X} \to \tilde{Y}) = \sum_{\mathcal{X}^{k+1}, \mathcal{Y}^{k+1}} \hat{P}_{\tilde{X}_{-k}^0 \tilde{Y}_{-k}^0, n}(a_0^k, b_0^k) \quad \times$$
$$\log \frac{\hat{P}_{\tilde{X}_{-k}^0 \tilde{Y}_0 | \tilde{Y}_{-k}^{-1}, n}(a_0^k, b_k | b_0^{k-1})}{\hat{P}_{\tilde{X}_{-k}^0 | \tilde{Y}_{-k}^{-1}, n}(a_0^k | b_0^{k-1}) \hat{P}_{\tilde{Y}_0 | \tilde{Y}_{-k}^{-1}, n}(b_k | b_0^{k-1})} \quad , \tag{7}$$

where conditional distributions are analogously found.

Given the DI rate estimate, it is now possible to test for causality. Assume that $\{(\tilde{X}_n, \tilde{Y}_n)\}$ is an irreducible and aperiodic Markov chain with memory length $k \geq 1$, with an all positive transition matrix. Assume also that the univariate process $\{\tilde{Y}_n\}$ is also a Markov chain with memory length k. Then, according to Kontoyiannis and Skoularidou [15] (theorem 6), if the process $\{\tilde{X}_n\}$ does not have a causal influence on $\{\tilde{Y}_n\}$, then, as the observation time of the sequences increases, the DI estimate converges in distribution to a χ^2-distribution, with $D = |\mathcal{Y}|^k(|\mathcal{X}|^{k+1} - 1)(|\mathcal{Y}| - 1)$ degrees of freedom, that is:

$$2n\hat{I}_n^{(k)}(\tilde{X} \to \tilde{Y}) \xrightarrow{\mathcal{D}} \chi^2(D). \tag{8}$$

Thus, the weighted DI rate estimate $2n\hat{I}_n^{(k)}(\tilde{X} \to \tilde{Y})$ is the test statistic. The objective is to test the null hypothesis H_0: $I_\infty(\tilde{X} \to \tilde{Y}) = 0$ (there is no causality) against the alternative hypothesis H_1: $I_\infty(\tilde{X} \to \tilde{Y}) > 0$ (there is causality). After the quantization step followed by the plug-in estimate, we accept H_0 if and only if $2n\hat{I}_n^{(k)}(\tilde{X} \to \tilde{Y}) < p$, where p is the 95 percentile of the χ^2-distribution with D degrees of freedom.

Moreover, we have separated the results of the tests on the trials according to 4 classes: true positive (TP), true negative

(TN), false positive (FP) and false negative (FN). A true positive happened if the test detected causality and there was an underlying causality between $\{X_n\}$ and $\{Y_n\}$ indeed. A false positive happened if the test detected causality but there was *not* an underlying causality. Similarly, a true negative happened if the test did not detect causality and there was not an underlying causality, and a false negative happened if the test did not detect causality but there was an underlying causality. With this classification of the trials, we were able to evaluate accuracy, sensitivity and specificity of the whole process of detecting causality.

In this paper, accuracy measures the global classification rate and is given by [22][23]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{9}$$

Sensitivity measures the ability of identifying causality when it actually exists:

$$Sensitivity = \frac{TP}{TP + FN}, \tag{10}$$

and specificity measures the ability of rejecting causality when it actually does not exist:

$$Specificity = \frac{TN}{TN + FP}. \tag{11}$$

V. Results

In this Section, we have evaluated the approach of firstly quantizing the continuous-state process with the three cited methods, plug-in estimating DI rate and then testing for causality. In order to do so, we have used three examples.

A. First example: simple delay

In the first example, there are two cases. In the first case, there is no underlying causality and we have two i.i.d. Gaussian processes, so $I_\infty(X \to Y) = 0$. In the second case, there is an underlying causality. The "driving" process $\{X_n\}$ is an i.i.d. Gaussian random process, with zero mean and unit variance, while the "response" process $\{Y_n\}$ is built according to:

$$Y_i = X_{i-1} + Z_i \tag{12}$$

where Z_n is also an i.i.d. Gaussian random process with zero mean and unit variance. As evaluated in our previous work [14], in the second case the analytical DI rate was $I_\infty(X \to Y) = 0.5$ bit.

We have quantized $\{X_n\}$ and $\{Y_n\}$ with the three different methods cited. We also used two sizes of quantization alphabets: $|\mathcal{X}| = |\mathcal{Y}| = 2$ or $|\mathcal{X}| = |\mathcal{Y}| = 6$. For each method and case, we have posteriorly used the plug-in estimator with memory length $k = 1$ and ran 50 trials with the time duration $n = 10^5$. Table II indicates the means of the estimates with the different quantization methods and alphabet size 2 and Table III indicates the means of the estimates with the different quantization methods and alphabet size 6.

TABLE II
DI rate estimate means, in bits, using the size of quantization alphabets $|\mathcal{X}| = |\mathcal{Y}| = 2$.

Causality	yes	no
Equidistant	4.9×10^{-5}	4.5×10^{-5}
Equipopulated	5.0×10^{-5}	4.6×10^{-5}
Symbolic	1.3×10^{-2}	4.3×10^{-5}

TABLE III
DI rate estimate means, in bits, using the size of quantization alphabets $|\mathcal{X}| = |\mathcal{Y}| = 6$.

Causality	yes	no
Equidistant	3.3×10^{-3}	4.2×10^{-3}
Equipopulated	7.7×10^{-3}	7.5×10^{-3}
Symbolic	3.0×10^{-2}	1.5×10^{-3}

It is clear that the estimates obtained in Tables II and III are very low when compared to the analytical value of the DI rate (0.5 bit), a feature also noticed in our previous work [14]. Moreover, it is an established fact that mutual information of continuous variables is the supremum of the quantized mutual information, achieved only by the refinement of partitioning [17]. Since DI is a sum of conditional mutual information [6], an analogous behaviour could be expected. However, in the DI rate estimates means, the symbolic method provides a greater difference between the case where there was causality and the case where there was not.

With the purpose of observing whether this difference was meaningful or not, we have evaluated hypothesis tests and the values for accuracy, sensitivity and specificity, as described in Section IV. Tables IV and V show the evaluated measures of classification.

TABLE IV
Rates of classification (%) of the first example, causal × not causal, using the size of quantization alphabets $|\mathcal{X}| = |\mathcal{Y}| = 2$.

Quantization	Equidistant	Equipopulated	Symbolic
Accuracy	55	53	91
Sensitivity	28	28	100
Specificity	82	78	82

TABLE V
Rates of classification (%) of the first example, causal × not causal, using the size of quantization alphabets $|\mathcal{X}| = |\mathcal{Y}| = 6$.

Quantization	Equidistant	Equipopulated	Symbolic
Accuracy	50	50	100
Sensitivity	0	100	100
Specificity	100	0	100

It is noticeable that the symbolic method of quantization, followed by the plug-in estimation of DI rate and hypothesis testing, provided the best method for detecting causality.

B. Second example: linear first order causal system

In this second example we have also simulated two cases. In the first case, we have a linear first order causal system,

where the input is a hidden Markov model [24]. The hidden states $\{\tilde{S}_n\}$ have the following transition probabilities: $P_{\tilde{S}_i|\tilde{S}_{i-1}}(0|0) = 0.9$, $P_{\tilde{S}_i|\tilde{S}_{i-1}}(1|0) = 0.1$, $P_{\tilde{S}_i|\tilde{S}_{i-1}}(0|1) = 0.05$, $P_{\tilde{S}_i|\tilde{S}_{i-1}}(1|1) = 0.95$. The observed sequence contidioned on the value of the hidden state $\tilde{S}_i = s_i$ had a Gaussian distribution: $X_i \sim \mathcal{N}(s_i, 1)$.

The linear system had the following differential equation:

$$\tau\dot{y} + y = Kx, \tag{13}$$

where $\dot{y}$ denotes the first derivative of the output y, the input x is equal to a realization of the "driving process" $\{X_n\}$, and the realization of the "response" process $\{Y_n\}$ is equal to the output y. In our simulations, $K = 1$ and $\tau = 5$. The system was simulated for the time interval $[1, n]$ with null initial conditions. After the simulation, both the input and the output were sampled at the time instants $(1, 2, 3, \ldots, n)$. Figure 1 shows an example of realization of the processes $\{\tilde{S}_n\}$, $\{X_n\}$ and $\{Y_n\}$.

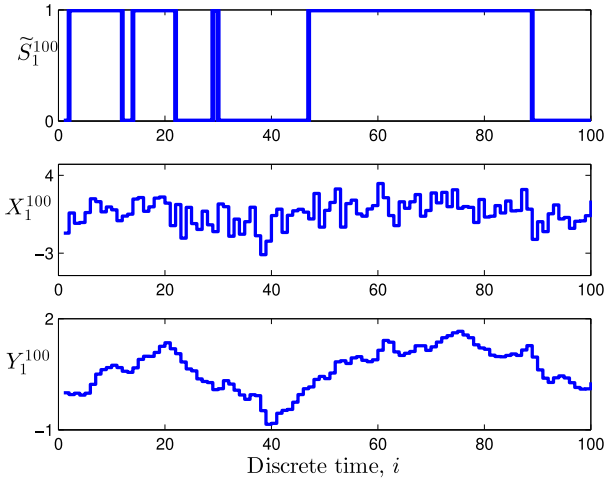

Fig. 1. Hidden states $\tilde{S}_1^{100}$, observable input X_1^{100} and the linear system output Y_1^{100}.

On the other hand, in the second case there was no underlying causality. We ran trials where both $\{X_n\}$ and $\{Y_n\}$ were independent hidden Markov models, that is: $X_i \sim \mathcal{N}(s_i, 1)$ and $Y_i \sim \mathcal{N}(e_i, 1)$, where $\{\tilde{S}_n\} \perp \{\tilde{E}_n\}$ are the hidden states with the same transition probabilities as previously stated.

With those examples, for each case and method we ran 50 trials of time duration $n = 10^5$. We have evaluated measures of accuracy, sensitivity and specificity of the capacity of the methods to detect causality, as before. Tables VI and VII present the results.

TABLE VI
RATES OF CLASSIFICATION (%) OF THE SECOND EXAMPLE, CAUSAL × NOT CAUSAL, USING THE SIZE OF QUANTIZATION ALPHABETS $|\mathcal{X}| = |\mathcal{Y}| = 2$.

Quantization	Equidistant	Equipopulated	Symbolic
Accuracy	89	92	92
Sensitivity	100	100	100
Specificity	78	84	84

TABLE VII
RATES OF CLASSIFICATION (%) OF THE SECOND EXAMPLE, CAUSAL × NOT CAUSAL, USING THE SIZE OF QUANTIZATION ALPHABETS $|\mathcal{X}| = |\mathcal{Y}| = 6$.

Quantization	Equidistant	Equipopulated	Symbolic
Accuracy	100	50	100
Sensitivity	100	100	100
Specificity	100	0	100

In this second example, the equidistant method of quantization had a similar performance to the symbolic method. However, with alphabet of size 2, the symbolic method had a better performance against the equidistant method. The equipopulated method had the worst performance with alphabet of size 6.

C. Third example: nonlinear causal relation

Similarly to the previous examples, in this third example we have two cases, one with an underlying causality and another with no causality. However, for this example we use a nonlinear relation. In the first case, we have the nonlinear causal relation:

$$Y_i = aY_{i-1} + b\cos(X_{i-1}) + \gamma Z_i, \tag{14}$$

where $\{X_n\}$ and $\{Z_n\}$ are independent, i.i.d. Gaussian processes with zero mean and unit variance ($X_i \sim \mathcal{N}(0, 1)$, $Z_i \sim \mathcal{N}(0, 1)$). We set $a = 0.8$, $b = 1$ and $\gamma = 0.5$. Figure 2 illustrates sample functions of the processes $\{X_n\}$ and $\{Y_n\}$ (which was built according to (14)), from time index 1 to 100.

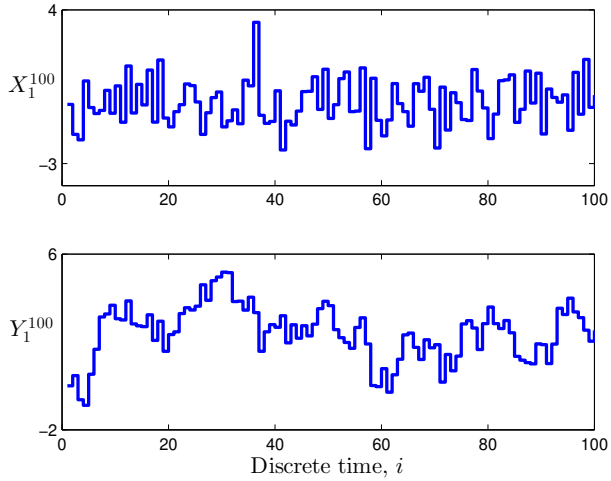

Fig. 2. Sample functions of the processes $\{X_n\}$ and $\{Y_n\}$.

In the second case, we have the same i.i.d. Gaussian $\{X_n\}$ process, but $\{Y_n\}$ is built as:

$$Y_i = aY_{i-1} + b\cos(W_i) + \gamma V_i, \tag{15}$$

where $\{W_n\}$ and $\{V_n\}$ are also independent, i.i.d. Gaussian processes with zero mean and unit variance. Again, we set $a = 0.8$, $b = 1$ and $\gamma = 0.5$.

Proceeding in an analogous way as in the previous examples, we have quantized the processes, evaluated DI rate plug-in estimates and performed statistical tests. Tables VIII and IX present the results.

TABLE VIII

RATES OF CLASSIFICATION (%) OF THE THIRD EXAMPLE, CAUSAL × NOT CAUSAL, USING THE SIZE OF QUANTIZATION ALPHABETS $|\mathcal{X}| = |\mathcal{Y}| = 2$.

Quantization	Equidistant	Equipopulated	Symbolic
Accuracy	84	47	87
Sensitivity	84	14	100
Specificity	84	80	74

TABLE IX

RATES OF CLASSIFICATION (%) OF THE THIRD EXAMPLE, CAUSAL × NOT CAUSAL, USING THE SIZE OF QUANTIZATION ALPHABETS $|\mathcal{X}| = |\mathcal{Y}| = 6$.

Quantization	Equidistant	Equipopulated	Symbolic
Accuracy	100	50	100
Sensitivity	100	100	100
Specificity	100	0	100

In this third example, the equidistant method had a comparable performance to the symbolic method. Equidistant method had a better specificity, while symbolic method had a better sensitivity (when the quantized processes had alphabet size 2). The equipopulated method had the worst performance again.

VI. CONCLUSION

In this paper we have investigated causality detection between two continuous-state discrete-time processes. This causality detection procedure consisted in quantization of the processes followed by DI rate plug-in estimation and hypothesis testing. The quantization was performed according to three different methods: equidistant, equipopulated and symbolic. In the majority of the simulated cases, the symbolic method had a performance equal to or better than the other methods, while the equipopulated method had the worst performance in causality detection. The only case where equidistant method slightly outperformed the symbolic method was in rejecting causality when it was actually abscent in a nonlinear causality relation (in the specificity measure in the third example, with alphabet size 2). Moreover, with the alphabet size 6 the symbolic method had a better performance than with size 2 (provided that there is a realization of the processes with a large time duration). Thus, the results suggest that the symbolic method of quantization, followed by plug-in DI rate estimation and hypothesis testing constitutes a simple but promising procedure for detecting causality between two continuous-state discrete-time random processes.

REFERENCES

[1] N. Wiener, "The theory of prediction," *Modern mathematics for engineers*, vol. 1, pp. 125–139, 1956.

[2] M. Wibral, R. Vicente, and J. T. Lizier, *Directed information measures in neuroscience*. Springer, 2014.

[3] A. E. Goodwell, P. Jiang, B. L. Ruddell, and P. Kumar, "Debates—does information theory provide a new paradigm for earth science? causality, interaction, and feedback," *Water Resources Research*, vol. 56, no. 2, p. e2019WR024940, 2020.

[4] R. G. James, N. Barnett, and J. P. Crutchfield, "Information flows? a critique of transfer entropies," *Physical review letters*, vol. 116, no. 23, p. 238701, 2016.

[5] T. Diamandis, Y. Murin, and A. Goldsmith, "Ranking causal influence of financial markets via directed information graphs," in *Conf. Rec. of CISS'2018*. IEEE, 2018, pp. 1–6.

[6] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *J Comput Neurosci*, vol. 30, pp. 17–44, 2011.

[7] J. M. de Assis and F. M. de Assis, "An application of directed information to infer synaptic connectivity," in *Conf. Rec. of SBrT'2016*, Santarém, Brazil, 2016, pp. 528–532.

[8] Z. Cai, C. L. Neveu, D. A. Baxter, J. H. Byrne, and B. Aazhang, "Inferring neuronal network functional connectivity with directed information," *Journal of neurophysiology*, vol. 118, no. 2, pp. 1055–1069, 2017.

[9] T. Tanaka, P. M. Esfahani, and S. K. Mitter, "Lqg control with minimum directed information: Semidefinite programming approach," *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 37–52, 2017.

[10] J. Jiao, H. H.Permuter, L. Zhao, Y.-H. Kim, and T. Weissman, "Universal estimation of directed information," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6220–6242, 2013.

[11] Y. Murin, J. Kim, and A. Goldsmith, "Tracking epileptic seizure activity via information theoretic graphs," in *Conf. Rec. of ACSSC'2016*. IEEE, 2016, pp. 583–587.

[12] W. Gao, S. Oh, and P. Viswanath, "Demystifying fixed k-nearest neighbor information estimators," *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5629–5661, 2018.

[13] C. Wang and M. M. Shanechi, "Estimating multiscale direct causality graphs in neural spike-field networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 857–866, 2019.

[14] J. M. de Assis and F. M. de Assis, "Estimation of directed information to processes assuming continuous values with ctw algorithm," in *Conf. Rec. of SBrT' 2017*, São Pedro, Brazil, 2017, pp. 528–532.

[15] I. Kontoyiannis and M. Skoularidou, "Estimating the directed information and testing for causality," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6053–6067, 2016.

[16] R. Malladi, G. Kalamangalam, N. Tandon, and B. Aazhang, "Identifying seizure onset zone from the causal connectivity inferred using directed information," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 7, pp. 1267–1283, 2016.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

[18] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, 1998.

[19] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.

[20] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical review letters*, vol. 88, no. 17, p. 174102, 2002.

[21] I. Grosse, "Applications of statistical physics and information theory to the analysis of dna sequences," Ph.D. dissertation, Boston University, 2000.

[22] M. O. Santos, J. M. de Assis, V. J. D. Vieira, and F. M. de Assis, "Ksg estimation of reconstruction delay to detect vocal disorders in nonlinear dynamical analysis," *Research on Biomedical Engineering*, vol. 34, no. 3, pp. 217–225, 2018.

[23] E. Alpaydin, *Introduction to Machine Learning*. The MIT Press, 2010.

[24] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.