SCIENTIFIC REP🅾RTS

OPEN

Causal Inference Based on the Analysis of Events of Relations for Non-stationary Variables

Yu Yin & Dezhong Yao

The main concept behind causality involves both statistical conditions and temporal relations. However, current approaches to causal inference, focusing on the probability vs. conditional probability contrast, are based on model functions or parametric estimation. These approaches are not appropriate when addressing non-stationary variables. In this work, we propose a causal inference approach based on the analysis of Events of Relations (CER). CER focuses on the temporal delay relation between cause and effect, and a binomial test is established to determine whether an "event of relation" with a non-zero delay is significantly different from one with zero delay. Because CER avoids parameter estimation of non-stationary variables *per se*, the method can be applied to both stationary and non-stationary signals.

Causality is a major concern in the study of complex systems; the detection and quantification of causal relations motivate many studies in various science domains. The foundation and connotation of causality have been continuously investigated for centuries by philosophers and scientists[1–4], and the main concepts are rooted in probabilistic theories[5]. A mathematically and statistically general definition of causality given by Wiener involves two facts[6]: 1. the statistical condition, where Y causes X, indicating that the corresponding probabilistic dependencies obtain, that is, $P(X) \neq P(X|Y)$, and 2. the temporal relations, where only the past and present may cause the future but the future cannot cause the past. However, it is difficult to apply this formula in practice for two reasons. First, it is difficult to obtain $P(X)$ without the effect of Y under natural conditions. This requires that experiments be conducted under highly controlled conditions. Therefore, we cannot determine whether the probability estimated from the time series X is $P(X)$ or $P(X|Y)$. Confidence is another issue, that is, whether or not $P(X|Y)$ is significantly different from $P(X)$ has to be considered.

Granger causality (GC)[7], which came from Wiener's definition of causality, and its derivatives[8–12] have been widely used in recent decades. GC used the prediction to infer the causality. The prediction accuracy of a joint model that includes "cause" and "effect" is better than a model of "effect" alone. Another familiar approach is transfer entropy (TE)[13], which depends on the overlap and dynamics of information transport by estimated probability distribution. However, it has been shown that TE and GC are equivalent for Gaussian variables[14], and asymptotically equivalent for general Markov models[15]. Also, information-theoretic measures often require substantially more data than regression methods such as GC[16]. Recently, Pearl suggested that the laws of probability theory do not dictate how one property of a distribution changes when another property is modified; therefore, causal mechanisms should be stated in frame of causal inference[17]. He proposed a structural causal model[18], which has received considerable attention in machine learning[19] and artificial intelligence[20]. Most importantly, all of these approaches focus on probabilistic dependencies. They may involve linear models or information theory, and the performance usually depends on the parameters estimation[21,22]. Furthermore, for widely existed non-stationary variables such as various physiological signals, the corresponding parameters estimation is still an open question[23,24]. To avoid complex parameter estimation, this work focuses on the second fact of causality, the temporal relations.

Following the causality definitions of Wiener and Pearl, we propose a new model of causality that "effect X" will occur after a certain lag τ when "cause Y" happens. Suppose that $F_\tau(X,Y)$ is joint probability distribution of X and Y, τ means Y at the earlier time. If Y doesn't cause X, F_τ will not be significantly different from others. In contrast, if Y causes X in a lag τ', the corresponding $F_{\tau'}$ would significantly differ from another F_τ such as F_0. In this model, we assume that the sampling rate of data is high enough so that τ' is not 0, because no information

Key Laboratory for NeuroInformation of Ministry of Education, Center for Information in Medicine, University of Electronic Science and Technology of China, Chengdu, 610054, China. Correspondence and requests for materials should be addressed to D.Y. (email: dyao@uestc.edu.cn)

transmission is instantaneous. Thus, F_τ is compared with F_0 to examine possible causality. To simplify the comparison, we suppose that X and Y are binary, only $\{X=1|Y=1\}$ is concerned, which means "cause Y" and "effect X" both happened. Simply take $M_\tau(\tau \geq 0)$ to note the event of relation (ER), which means that X and Y are observed, where Y is delayed by τ. $K_\tau(\tau \geq 0)$ is the number of ER during the observation, and $p(M_\tau)$ is the probability of those ER. Thus comparison is degenerated to a binomial test, and Poisson test (approximation of binomial test) is employed to simplify calculation (see Methods).

Results

To test the efficiency of the CER, our causal inference approach based on the analysis of ER $M_\tau(\tau \geq 0)$, the neuronal spikes interaction model[25] with delay was simulated. In basic two-node interaction system based on the neuronal spikes interaction model, each node is driven by an event probability p to generate the binary time series. When node $Y=1$, node X will be "1" with the interactive probability p_{xy} after a delay τ (schematic in Fig. 1a). The time series may be stationary or non-stationary depending on whether event probability p is constant or time-varying.

CER was tested for different types of lag and both stationary and non-stationary simulations. Figure 1b shows that K_τ of M_τ for different lags were not significantly different from K_0 when the two stationary nodes were assumed to be non-causal. As Fig. 1c depicts the delay was a uniform distribution from (0, 100], which means that the "effect" can occur at any time after the "cause" appears. And $p(M_\tau)$ was not higher than the upper limit, indicating that the Poisson test failed to reject the null hypothesis that $p(M_0)$ and $p(M_\tau)$ are equal. In practice, the delay between two events having a causal relation is smaller than a certain value. Thus if the delay follows an unusually wide uniform distribution, it might be logically assumed to be non-causal. In contrast, Fig. 1(d–i) illustrate that the CER detected all pre-designed causality at accurate delay times under other conditions.

Figure 2a illustrates a randomly generated (random walk) series of non-stationary probability p used to generate a binary time series to simulate non-stable spontaneous activity in one node. The non-stationarity of those time-varying p series was verified by Dickey-Fuller test ($P < 0.05$) in all simulations. Two moving windows of different length were employed to calculate entropy of the non-stationary binary time series, respectively (Fig. 2b). The results indicated that the moving window technique might not be efficient for non-stationary data.

Detection rate and false-positive rate are two critical issues in a causal inference test, and five possible outcomes might be encountered in practice: 1. reject the null hypothesis at the right time (strongly correct); 2. reject the null hypothesis in the right causal direction without a well-defined delay (weakly correct); 3. failure to reject the null hypothesis when the variables are causally related (missing detection); 4. reject the null hypothesis but the causal direction is wrong (error type I); and 5. reject the null hypothesis but the variables are not causally related (error type II). In our study the ratios of the five possible outcomes were calculated using 10000 runs of the numerical simulation and different simulation data were tested at three α levels (Fig. 3). The error type II was almost zero ($<0.7\%$ at α of 0.0005) in all of these simulations, which are thus not shown. As Fig. 3 displays, the CER dominantly pointed to "strongly correct" outcome. The missing detection cases occurred mainly for data with weak interaction. Therefore, the CER exhibited a good performance in terms of excluding non-causality data with few errors. The error type I occurred mainly at α level of 0.05. At a more stringent α of 0.005 or even 0.0005, the ratio of the error type I decreased to nearly zero, and therefore, we may choose a smaller α when applying the CER.

Specifically, the detection rate could still be 100%, even for a non-stationary and Gaussian-distributed delay case (Fig. 3d) if the size of the dataset was sufficiently large and if the interaction was not overly low. Under this condition, $p(Y_\tau X)$ may be significantly different from $p(Y_0 X)$ at more than one delay τ (Fig. 1i). Such type of data is a substantial challenge for hypothesis testing, and the CER maintains high performance in this case.

Now, we consider a possible complex case, directed acyclic graph (DAG). In DAG, nodes can be indirectly relevant to each other or respond to a common input[26]. The simulation of DAG was basically the same as the interaction model we previously used except that the node number was three instead of two. As the statistics we investigated are the temporal relations, inferring the direction of causality in DAG can be realized without knowledge of interested third-party. Therefore, the CER examined nodes in pair. It detected all pre-designed causality at an accurate delay time (Fig. 4).

Discussion

One merit of the CER is the ability to process non-linear and non-stationary variables because it is based on the statistic variable "ER", which does not depend on the dynamic process of variables *per se*. Moreover, causal inference may be greatly impacted by the performance of prediction models when GC and its derivatives are applied. In a complex system, the dynamic process of nodes may be totally different, and a joint model would be difficult to construct because of heterogeneous properties (e.g., being non-stationary or nonlinear). A copula approach[27,28] has recently been proposed to reveal nonlinear, non-stationary causality and deal with binary events[29,30]; however, the algorithm requires kernel estimation, and it is still a challenge.

The core of the CER is the temporal relations in the definition of causality. Temporal relations were also implicated in other approaches[31]. In the widely applied GC, they are embedded in using the order of the regressive model. Differently, Pearl introduced "$do(x)$" for setting $X=x$ and called the mapping from x to $P(y|do(x))$ for all x the causal effect of X on Y[17]. The operator "$do(x)$" is used to emphasize the occurrence of "x". The "x" that previously occurred changes the distribution of "Y", which implies temporal relations in causality. But in most conditions the above-mentioned approaches only identify past and future. How long will the "effect" delay after "cause" occurs? Or what does a distribution of the lags follow? Such information would also be meaningful for causality inference. In the CER, the temporal relation is a crucial variable. Our approach may detect not only the causal relation but also the lag or lag distribution.
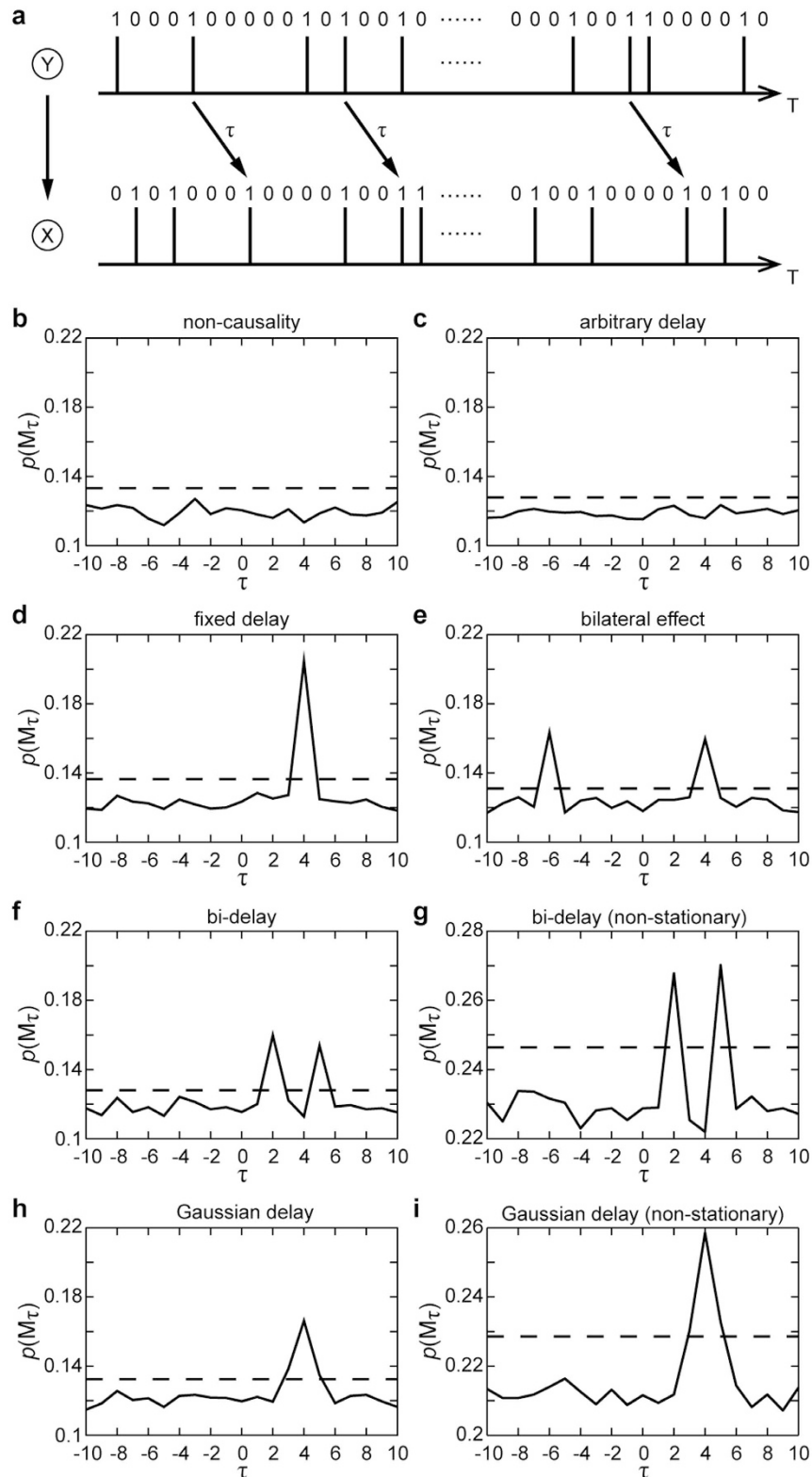
Figure 1. CER in discrete binary models. (**a**) Schematic of the causality of a two-node binary system. In following simulations, the event probability of the stationary series is fixed at 0.3; interactive probability $p_{YX} = 0.6$, except as specifically stated. (**b**) Non-causality: $p_{YX} = 0$, (**c**) Arbitrary delay: τ follows a uniform distribution ($0 < \tau \leq 100$). (**d**) Fixed delay: $\tau = 4$. (**e**) Bilateral effect: $p_{YX} = 0.3$, $\tau = 4$; $p_{YX} = 0.3$, $\tau = 6$; no reciprocal action. (**f**) Bi-delay: $\tau = \{2,5\}$ with the same probability 0.5. (**g**) Bi-delay in non-stationary series. (**h**) Gaussian delay: using $\tau = \{3, 4, 5\}$ with probabilities $\{0.225, 0.55, 0.225\}$, respectively, to simulate a Gaussian distribution. (**i**) Gaussian delay in non-stationary series. The horizontal axis shows the delay time, and the vertical axis shows the probability (frequency) of ER. The dotted line is the upper limit by delay 0 at $\alpha = 0.005$. The length of the simulation dataset is 5,000 points.
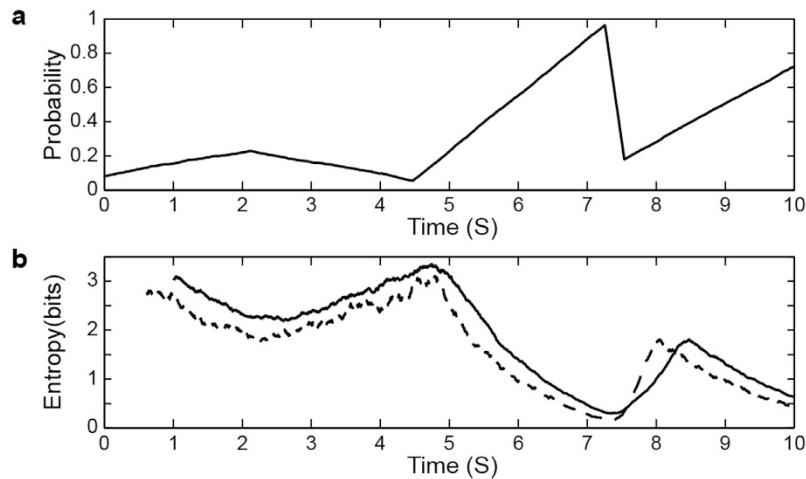
Figure 2. An example of non-stationary data. (**a**) Time-varying event probability; (**b**) Time-varying entropy of event series calculated with two different moving window lengths. Solid line: window length 1000 points, dotted line: window length 500 points. The data length is 10 seconds with a sampling rate of 1000 per sec.

In general, a high sample rate allows temporal properties to be exposed in the data and utilized in causality. If the temporal resolution is low, which means that "cause" and "effect" occur at the same sampling time according to Eq. (7), K_0 will be larger than all K_τ and unable to determine the direction when nodes are causally related. However, one can still determine whether causality exists because the sign of the inequality is true in statistics.

It is common to record discrete data in many studies. Although the values of variables are often numerous, not all of the values are important. In many practical cases, binary or multiple values are common. In addition, discrete events are often objective reflections of many phenomena, and defining a discrete event is typically a goal during data processing. Moreover, data discretization provides information to answer particular questions. For example, in the analysis of relations between the prices (continuous value) of stocks *A* and *B*, it is informative to identify whether the price variations (binary value) of stock *A* affect the price variations of stock *B* during a period of time.

In summary, the CER approach features temporal relations, one crucial aspect of causality, and uses them as the basis of causal inference. It can be applied to non-linear, non-stationary, and binary variables. In addition, the CER can also be extended to systems of multiple values to address other causal problems. In the future use of the approach, the third-party variables may be included to reduce error type II. An easy-to-use Matlab tool about causal inference can be downloaded from http://www.neuro.uestc.edu.cn/CER.html.

Methods

Let us briefly review the causality definition summarized by Granger[32]. "For ease of exposition, a universe is considered where all variables are measured just at prespecified time points at constant intervals $t = 1, 2\ldots$ When at time *n*, let all of the knowledge in the universe available at that time be denoted Ω_n and denote by $\Omega_n - Y_n$ this information except the values taken by a variable Y_t up to time *n*, where $Y_n \in \Omega_n$. Suppose that we are interested in the proposition that the variable *Y* causes the variable *X*. At time *n*, the value X_{n+1} will be a random variable and so can be characterized by probability statements Prob $(X_{n+1} \notin A)$ for a set *A*". If Y_n causes X_{n+1}, the following general definition can be suggested:

$$\text{Prob}(X_{n+1} \notin A|\Omega_n) \neq \text{Prob}(X_{n+1} \notin A|\Omega_n - Y_n) \quad \text{for some } A \tag{1}$$

In this equation, the temporal relations are illustrated by a general delay indicated by the subscript "$n + 1$". Here "$+1$" just means an interval to distinguish present between future.

But in the real world, the lag between "cause" and "effect" must be within a limited period. It might be a probability distribution or several discrete points. However, it would not be equal to zero because information transmission is not instantaneous. Therefore, the causality may be evaluated using the differences between a delay of zero and other options. We proposed that an important factor of causality is implicit stable temporal relations between variables. An appropriate statistical test and statistical variable are needed to confirm these differences. For simplicity, we assume that time series are binary "stochastic events", i.e., the value of the variable is "1" when events have occurred and "0" when they have not occurred. We used *p* to denote the probability of the "stochastic event".

In discrete cases, we can rename "stochastic variable" as "event". Let us place events *X* and *Y* in the context of all of the knowledge in the universe; let Ω be a sample space of all measurable elements of interest in the universe; let *F* be a σ-field (collection of subsets of the Ω); and let *p* be a probability that is defined in measurable space (Ω, F), with $p(\Omega) = 1$. The triple (Ω, F, p) is called a probability space, and events *X* and *Y* belong to the set *F*. The subscript *n* in Eq. (1) means that the past and present can cause the future, but the future cannot cause the past[31]. In time series, the time factor of causality is denoted by $\tau(\tau > 0)$. Here, we employed time delay and binarization to depict event *Y* that previously occurred. As a result, for discrete events, Eq. (1) is still valid, and we can adapt it to:

$$p(X_{t+\tau}) \neq p(X_{t+\tau}|\overline{Y_t}) \tag{2}$$
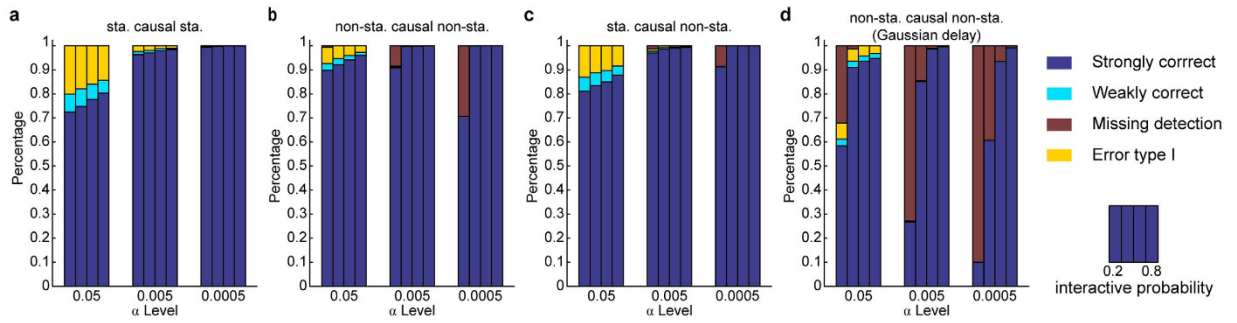
Figure 3. CER computational stability of a two-node system with different interaction probabilities at different test levels α. In each sub-graph, the statistical test at three α levels of 0.05, 0.005 and 0.0005 are shown in three bar graph groups. For each α level, the four bars are the results of using four interaction probabilities, 0.2, 0.4, 0.6 and 0.8. The total parameters are the same as in Fig. 1. Fixed delay (**a–c**): $\tau = 4$. Gaussian delay (**d**): $\tau = \{2, 3, 4, 5, 6\}$ with probabilities $\{0.015, 0.21, 0.55, 0.21, 0.015\}$, respectively. The sta. denotes stationary variables, and non-sta. denotes non-stationary variables.
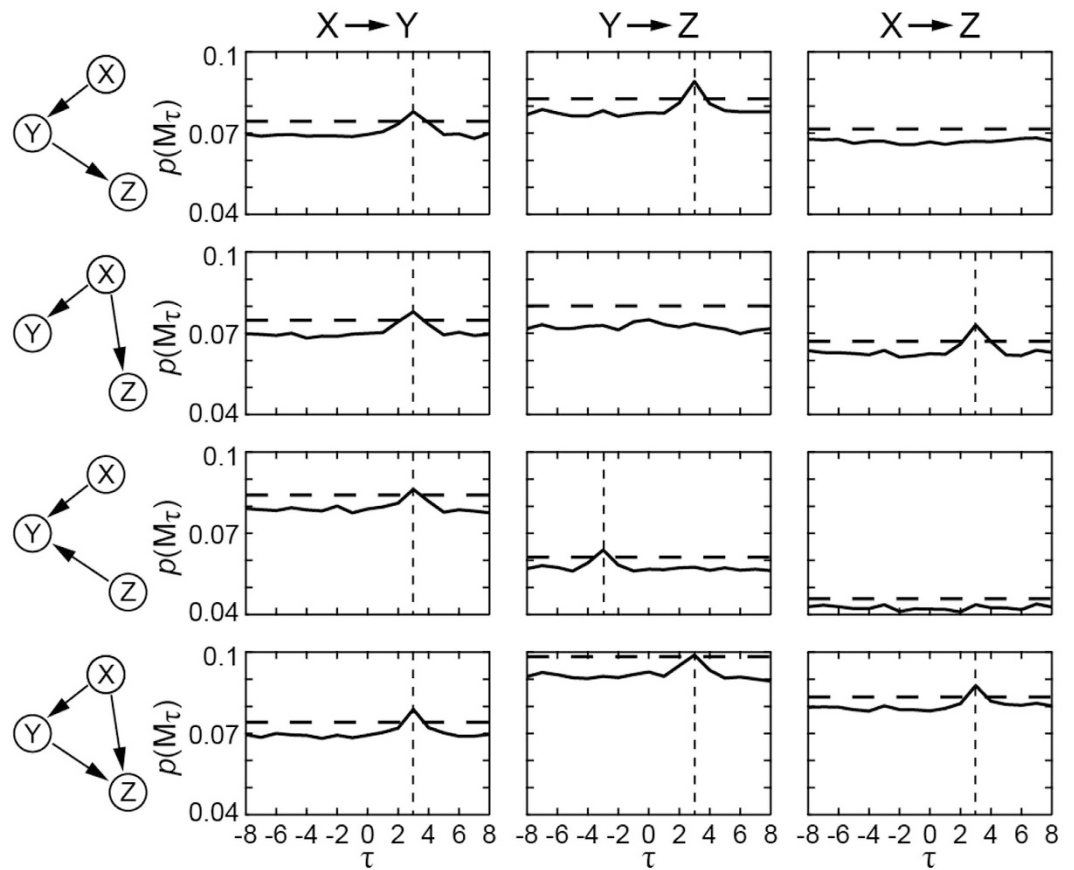
Figure 4. CER in DAG models of three nodes. The first column shows interaction of three nodes. The other columns are CER for different node pairs. The event probability is non-stationary; total interactive probability $p = 0.6$. Gaussian delay: $\tau = \{2, 3, 4\}$ with probabilities $\{0.225, 0.55, 0.225\}$, respectively. The horizontal axis shows the delay time, and the vertical axis shows the probability (frequency) of ER. The dotted line is the upper limit by delay 0 at $\alpha = 0.005$. Note that there is a negative delay time because the direction of causality is tested for node Y to Z in the third column.

Here, $\overline{Y}$ is the complementary set of Y. Event series Y shifted by τ along the timeline, denoted by Y_τ. From here on, the subscript t is omitted for it is common for both X and Y. Such an operation does not change the probability of the occurrence of the event Y, i.e., $p(Y) = p(Y_\tau)$. Then we adapt Eq. (2) to

$$p(X) \neq p(X|\overline{Y}_\tau)$$

$$p(X) \neq \frac{p(\overline{Y}_\tau|X)p(X)}{p(\overline{Y}_\tau)}$$

$$1 - p(Y_\tau) \neq p(\overline{Y}_\tau|X)$$

Because events X and Y belong to the set F, we have $p(\overline{Y}_\tau|X) = 1 - p(Y_\tau|X)$; then,

$$1 - p(Y_\tau) \neq 1 - p(Y_\tau|X)$$

$$p(Y_\tau|X) \neq p(Y_\tau)$$

$$p(Y_\tau X) \neq p(Y_\tau)p(X)$$

$$p(Y_\tau X) \neq p(Y)p(X) \tag{3}$$

Since no information transmission is instantaneous, the τ is not equal to zero. If $\tau = 0$ event Y does not cause event X. Thus Eq. (2) can be edited as follows:

$$p(X) = p(X|\overline{Y}_0) \tag{4}$$

Following the same derivation from Eq. (2) to Eq. (3); from Eq. (4), we have

$$p(Y_0 X) = p(Y)p(X) \tag{5}$$

Next, merge Eq. (3) and Eq. (5) as follows:

$$p(Y_0 X) \neq p(Y_\tau X) \tag{6}$$

Eq. (6) illustrates that the joint probability $p(XY)$ of X and Y, observed without delay, is different from that of X and Y_τ if Y causes X with delay τ. Conversely, if X and Y have no causal relation, then the joint probabilities $p(Y_0 X)$ and $p(Y_\tau X)$ will not be significantly different regardless of the value of τ. This means that the change in the observation time of one event does not affect the joint probability *per se* or any other non-causal event. Let us assume that someone rolls a coin first and a dice later. The non-causality of the experiments will not be changed if the person rolls a dice in one hour or even tomorrow. Hence, if the relation between events X and Y is not causal, then events X and Y_τ are not causal.

Now, let us denote the joint event of Y_τ and X as a new statistical event M_τ, event of relation (ER), which means that events X and Y are observed, where Y is delayed by τ. As a result, Eq. (6) can be rewritten in terms of event M:

$$p(M_0) \neq p(M_\tau) \quad (\tau > 0) \tag{7}$$

In this equation, "Y and X" are considered to be one object M, which is decided by the relations of events X and Y and not by X or Y *per se*. Hence, it is no longer necessary to know the probability of X or Y. Even if X and Y are non-stationary variables, the causal inference of ER can still be evaluated because it is investigated by the new statistical variable ER instead of the original variables.

In Eq. (7), $p(M_0)$ is compared to $p(M_\tau)$ with different delays to determine whether the sign of the inequality is true. For binary series X and Y, only $\{X = 1|Y = 1\}$ is concerned, which means "cause Y" and "effect X" both happened; thus the ER values are 1 (cause and effect both happened) and 0 (other conditions). Therefore, a binomial test can be employed. The null hypothesis of Eq. (7) is that two probabilities are equal likely to occur. $p(M_0)$ is assumed to be the expected probability under the condition of non-causality. If the frequency of the ER observed with delay τ is significantly higher or lower than $p(M_0)$, then causality existed between the series data X and Y.

In the computation, we adopted a Poisson distribution, which is an effective approximation of a binomial distribution and easy to compute if the dataset is sufficiently large. When the number of observations is n, event M_0 occurs K_0 times, and event M_τ occurs K_τ times. If $Poission(K_\tau, K_0) > 1 - \alpha$, K_τ is believed to be significantly greater than K_0 at the α significance level. In general, the "cause" is considered to increase the probability of the "effect". If the "cause" decreases the probability of the "effect", then it is called negative causality. Under this condition, $\alpha < Poisson(K_\tau, K_0)$ can be used to determine whether K_τ is significantly smaller than K_0. Apparently, both increase and decrease are processed in the same manner, except for the hypothesis test (i.e., right tail, left tail or two-tailed, respectively for increase, decrease or unknown cases). In this paper, only the increase case is considered in the simulations.

The algorithm of CER for binary series is summarized as follows:

1. Calculate the number of ERs with delay 0, $K_0 = sum\{X = 1|Y = 1\}$;
2. Calculate the numbers of ERs with different delay τ, $K_\tau = sum\{X = 1|Y = 1\}$; Y_τ means series Y delayed by τ;
3. The Poisson cumulative distribution function of ER is given by

$$p = F(x|K_0) = e^{-k_0} \sum_{i=0}^{floor(x)} \frac{K_0^i}{i!};$$

4. If $p_\tau = F(K_\tau|K_0) < \alpha$ indicates that K_τ is significantly smaller than K_0; then Y has negative causality with X after delay τ;

5. If $p_\tau = F(K_\tau|K_0) > 1 - \alpha$ indicates that K_τ is significantly bigger than K_0; then Y has positive causality X after delay τ.

References

1. Russell, B. On the notion of cause. *Proc Aristot Soc.* **13,** 1–26, (1912).
2. Reichenbach, H. *The direction of time* (Univ. of California Press, 1956).
3. Good, I. J. A theory of causality. *Br J Philos Sci.* **IX 36,** 307–310 (1959).
4. Suppes, P. *A probabilistic theory of causality* (North-Holland, 1970).
5. Cox, D. R. & Wermuth, N. Causality: A statistical view. *Int Stat Rev.* **72,** 285–305 (2004).
6. Wiener, N. The theory of prediction, In *Modern Mathematics for Engineers* Vol. 1. (ed. Beckenbach, E. F.) 125–139 (New York: McGraw-Hill, 1956).
7. Granger, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica J Econ Soc.* **37(3),** 424–438 (1969).
8. Kamiński, M., Ding, M., Truccolo, W. A. & Bressler, S. L. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol Cybern.* **85,** 145–157 (2001).
9. Marinazzo, D., Pellicoro, M. & Stramaglia, S. Kernel method for nonlinear Granger causality. *Phys Rev Lett.* **100,** 144103 (2008).
10. Sommerlade, L. *et al.* Inference of Granger causal time-dependent influences in noisy multivariate time series. *J. Neurosci Meth.* **203,** 173–185 (2012).
11. Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Sci.* **338,** 496–500 (2012).
12. Wang, X., Chen, Y., Bressler, S. L. & Ding, M. Granger causality between multiple interdependent neurobiological time series: blockwise versus pairwise methods. *Int J Neural Syst.* **17(02),** 71–78 (2007).
13. Schreiber, T. Measuring information transfer. *Phys Rev Lett.* **85,** 461–464 (2000).
14. Barnett, L., Barrett, A. B. & Seth, A. K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys Rev Lett.* **103,** 238701 (2009).
15. Barnett, L. & Bossomaier, T. Transfer entropy as a log-likelihood ratio. *Phys Rev Lett.* **109(13),** 138105 (2012).
16. Pereda, E., Quiroga, R. Q. & Bhattacharya J. Nonlinear multivariate analysis of neurophysiological signals. *Prog Neurobiol.* **77(1),** 1–37 (2005).
17. Pearl, J. Causal inference in statistics: An overview. *Stat Surv.* **3,** 96–146 (2009).
18. Pearl, J. *Causality: models, reasoning, and inference* (Cambridge University Press, 2000).
19. Andrieu, C., De Freitas, N., Doucet, A. & Jordan, M. I. An introduction to MCMC for machine learning. *Mach Learn.* **50,** 5–43 (2003).
20. Nilsson, N. J. *Artificial Intelligence: A New Synthesis* (Elsevier, 1998).
21. Lindner, M., Vicente, R., Priesemann, V. & Wibral, M. TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci.* **12(1),** 119 (2011).
22. Dhaene, G. & Jochmans, K. An adjusted profile likelihood for non-stationary panel data models with fixed effects. Working paper, KU Leuven (2012).
23. Möller, E., Schack, B., Arnold, M. & Witte, H. Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models. *J Neurosci Meth.* **105(2),** 143–158 (2001).
24. Ding, M., Bressler, S. L., Yang, W. & Liang, H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biol Cybern.* **83(1),** 35–45 (2000).
25. Quinn, C. J., Coleman, T. P., Kiyavash, N. & Hatsopoulos, N. G. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *J Comput Neurosci.* **30,** 17–44 (2011).
26. Janzing, D., Balduzzi, D., Grosse-Wentrup, M. & Schölkopf, B. Quantifying causal influences. *Ann Statist.* **41,** 2324–2358 (2013).
27. Hu, M. & Liang, H. A copula approach to assessing Granger causality. *NeuroImage.* **100,** 125–134 (2014).
28. Hu, M. *et al.* Copula Regression Analysis of Simultaneously Recorded Frontal Eye Field and Inferotemporal Spiking Activity during Object-Based Working Memory. *J Neurosci.* **35(23),** 8745–8757 (2015).
29. Kim, S., Putrino, D., Ghosh, S. & Brown, E. N. A Granger Causality Measure for Point Process Models of Ensemble Neural Spiking Activity. *PLoS Comput Biol.* **7(3),** e1001110 (2011).
30. Hu, M., Li, W. & Liang, H. A Copula-based Granger Causality Measure for the Analysis of Neural Spike Train Data. *IEEE ACM T Comput Bi.* **PP(99),** 1–1 (2014).
31. Kleinberg, S. & Hripcsak, G. A review of causal inference for biomedical informatics. *J Biomed Inform.* **44,** 1102–1112 (2011).
32. Granger, C. W. J. Testing for causality: a personal viewpoint. *J Econ Dynam Control.* **2,** 329–352 (1980).

Acknowledgements

Author Contributions

Y.Y. conceived and conducted the method, D.Y. analyzed results. Y.Y. and D.Y. wrote the paper. All authors reviewed the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Yin, Y. and Yao, D. Causal Inference Based on the Analysis of Events of Relations for Non-stationary Variables. *Sci. Rep.* **6**, 29192; doi: 10.1038/srep29192 (2016).