

Exploratory time series causality (working title)

June 18, 2015

Contents

1	Introduction	1
2	Causality Studies	1
2.1	Foundational Causality	1
2.1.1	Philosophical Studies	1
2.1.2	Natural Science Studies	2
2.1.3	Psychological (and other Social Science) Studies	2
2.2	Phenomenal Causality	2
2.2.1	Statistical Causality	2
2.2.2	Data Causality	2
2.2.3	Time Series Causality	2
3	Time Series Causality	2
3.1	Granger Causality	2
3.2	Transfer Entropy	2
3.2.1	Background	2
3.2.2	Practical Usage	3
3.3	State Space Reconstruction Causality	3
3.4	Lagged Cross-Correlation	3
3.5	Penchants and Leanings	3
4	Exploratory Causal Analysis	3
4.1	Synthetic data examples	4
4.2	Empirical data example	4
5	Conclusions	4

1 Introduction

Why is this work needed, where does it fit in the grand scheme of causality studies, what exactly is the goal, what is not the goal, etc.?

2 Causality Studies

Literature review. This taxonomy will not be complete, nor will the boundaries be clean. Several authors, particularly in theoretical physics, blur boundaries, e.g., between science and philosophy.

2.1 Foundational Causality

2.1.1 Philosophical Studies

Aristotle, Hume, . . . , Good, Suppes

2.1.2 Natural Science Studies

causality in physics, etc.

2.1.3 Psychological (and other Social Science) Studies

perceptions of causality

2.2 Phenomenal Causality

2.2.1 Statistical Causality

Fisher, Dawid, Rubin, Pearson, ...

2.2.2 Data Causality

Pearl, Kleinberg, ...

2.2.3 Time Series Causality

This is just an explanation of where these techniques fit in the taxonomy. In-depth descriptions are the next chapter.

3 Time Series Causality

Introduction and (brief) exploration of the five main classes of times series causality techniques

3.1 Granger Causality

3.2 Transfer Entropy

In 2000, Schreiber introduced transfer entropy as “an information theoretic measure is derived that quantifies the statistical coherence between systems evolving in time” [9]. The quantity has become one of the primary tools of information theoretic times series causality [4, 6], which focuses on using entropies and mutual information for causal inference between times series data.

3.2.1 Background

Consider a random variable \mathbf{X} that takes value X_n with probability p_n with $n = 1, 2, \dots, N_X$. The probability distribution $P(\mathbf{X} = X_n) = p_n \forall X_n$ has a discrete Shannon entropy, H_X defined as

$$H_X = - \sum_{n=1}^{N_X} p_n \log_2 p_n \quad , \quad (1)$$

where the logarithm base determines the entropy units, which, in this case, is “bits” [10, 9, 4, 6]. The Shannon entropy was developed as a measure of “of how uncertain we are of the outcome” $\mathbf{X} = X_n$ [10].

The error made by incorrectly assuming $P(\mathbf{X} = X_n) = q_n$ (rather than p_n) is the Kullback entropy¹

$$K_X = \sum_{n=1}^{N_X} p_n \log_2 \frac{p_n}{q_n} \quad , \quad (2)$$

where, again, the base of the logarithm is due to the unit choice [7, 9, 4, 6]. Henceforth throughout this subsection, the logarithm base notation will be dropped and all logarithms should be assumed base 2 (i.e., everything is in units of bits) unless otherwise noted.

¹This quantity is also known as the Kullback-Leibler divergence, relative entropy, and discrimination information, among others. Kullback noted in 1987 that this quantity could be found under nine different names in the literature [3].

Consider a second random variable \mathbf{Y} that takes value Y_m with probability p_m with $m = 1, 2, \dots, N_Y$. The joint entropy is defined as

$$H_{X,Y} = - \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} p_{n,m} \log p_{n,m} , \quad (3)$$

where $p_{n,m}$ is the joint probability $P(\mathbf{X} = X_n, \mathbf{Y} = Y_m)$. If the two random variables are statistically independent, then $H_{X,Y} = H_X + H_Y$ [4], which is motivation for the introduction of the mutual information $I_{X,Y}$ as

$$I_{X,Y} = H_X + H_Y - H_{X,Y} = \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} p_{n,m} \log \frac{p_{n,m}}{p_n p_m} , \quad (4)$$

where the last equality can also be seen as the Kullback entropy due to assuming $P(\mathbf{X} = X_n, \mathbf{Y} = Y_m) = p_n p_m$ [6]. The mutual information is symmetric and is, therefore, not much use as a time series causality tool. Time lags and conditional entropies, i.e., $H_{X|Y} = H_{X,Y} - H_Y$, can be used to make the mutual information non-symmetric, but such modifications are not easily interpreted in terms of information flow [9] (the assumption is that a flow of information from one time series to another may be indicative of a driving relationship).

Schreiber's approach for using entropies to study dynamical structure of data was to focus on transitional probabilities in the system [9]. Suppose at time t , $\mathbf{X}(t) = X_n$ with some probability $P(\mathbf{X}(t) = X_n^{(t)}) = p_n$. The additional temporal structure allows for more complicated questions such as "what is the uncertainty in $\mathbf{X}(t)$ given some value of $\mathbf{X}(t-1)$?" The more basic question is how to define the transitional probabilities themselves, i.e., what is $P(\mathbf{X}(t) = X_n | \mathbf{X}(t-1) = X_{n-1}, \mathbf{X}(t-2) = X_{n-2}, \dots, \mathbf{X}(0) = X_0) = p_{n|n-1,n-2,\dots,0}$?

Assume $\mathbf{X}(t)$ is governed by a Markov process; i.e., $p_{n|n-1,n-2,\dots,0} = p_{n|n-1}$ [?]. The basic concept of transfer entropy is to measure the error made by assuming the Markov process generating \mathbf{X} does not depend at all on the second time series \mathbf{Y} . This error is quantified, in analogy to the mutual information expression shown above, with the Kullback entropy as

$$T_{Y \rightarrow X} = \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} p_{n+1,n,m} \log \frac{p_{n+1|n,m}}{p_{n+1|n}} , \quad (5)$$

where $p_{n+1,n,m} = P(\mathbf{X}(t+1) = X_{n+1}, \mathbf{X}(t) = X_n, \mathbf{Y}(\tau) = Y_m)$ is the joint probability of \mathbf{X} at times t and $t+1$ with \mathbf{Y} at time τ , and the two conditional probabilities are $p_{n+1|n,m} = P(\mathbf{X}(t+1) = X_{n+1} | \mathbf{X}(t) = X_n, \mathbf{Y}(\tau) = Y_m)$ and $p_{n+1|n} = P(\mathbf{X}(t+1) = X_{n+1} | \mathbf{X}(t) = X_n)$. Schreiber notes that the "most natural choices" for τ are $\tau = t$ and $\tau = 0$ [9]. The quantity $T_{Y \rightarrow X}$ is called the transfer entropy. If $H_{X(t+1)|X(t)}$ is the conditional Shannon entropy at time $t+1$ given t , then the above expression can be rewritten as $T_{Y \rightarrow X} = H_{X(t+1)|X(t)} - H_{X(t+1)|X(t), Y(\tau)}$ [6].

Transfer entropy is often considered just one part of an information-theoretic approach to time series causality [5], but it will be the primary information-theoretic time series causality tool used in this work.

3.2.2 Practical Usage

It can be shown that the transfer entropy is equivalent to Granger causality (up to a factor of 2) if \mathbf{X} and \mathbf{Y} are jointly multivariate Gaussian (i.e., both variables follow normal (Gaussian) distributions individually and jointly) [2]. As Barnett et al. point out in their paper, this result provided "for the first time a unified framework for data-driven causal inference that bridges information-theoretic and autoregressive methods." The Gaussian restriction of the result implies the use of Granger causality and transfer entropy for exploratory causal analysis, in general, still depends on the data under consideration. Other authors have also explored to relationship between Granger causality and information-theoretic causality measures such as transfer entropy; e.g., see [1, 8].

3.3 State Space Reconstruction Causality

3.4 Lagged Cross-Correlation

3.5 Penchants and Leanings

4 Exploratory Causal Analysis

Applying the time series causality tools

4.1 Synthetic data examples

4.2 Empirical data example

5 Conclusions

Future work, etc.

References

- [1] Pierre-Olivier Amblard and Olivier J. J. Michel. The relation between granger causality and directed information theory: A review. *Entropy*, 15(1):113, 2012.
- [2] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett.*, 103:238701, Dec 2009.
- [3] Subir Ghosh, Kenneth P. Burnham, Nico F. Laubscher, Gerard E. Dallal, Leland Wilkinson, Donald F. Morrison, Milton W. Loyer, Bennett Eisenberg, Solomon Kullback, Ian T. Jolliffe, and Jeffrey S. Simonoff. Letters to the editor. *The American Statistician*, 41(4):pp. 338–341, 1987.
- [4] Katerina Hlavkov-Schindler, Milan Palu, Martin Vejmelka, and Joydeep Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1 – 46, 2007.
- [5] Katerina Hlavkov-Schindler, Milan Palu, Martin Vejmelka, and Joydeep Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1 – 46, 2007.
- [6] A. Kaiser and T. Schreiber. Information transfer in continuous processes. *Physica D: Nonlinear Phenomena*, 166(12):43 – 62, 2002.
- [7] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.
- [8] M. LUNGARELLA, K. ISHIGURO, Y. KUNIYOSHI, and N. OTSU. Methods for quantifying the causal structure of bivariate time series. *International Journal of Bifurcation and Chaos*, 17(03):903–921, 2007.
- [9] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.
- [10] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal, The*, 27(3):379–423, July 1948.